University of Venda

MEDIUM TERM LOAD FORECASTING IN SOUTH AFRICA USING GENERALIZED ADDITIVE MODELS WITH TENSOR PRODUCT INTERACTIONS

By

Ravele Thakhani

11630148

Dissertation submitted for Masters of Science degree in Statistics

in the

Department of Statistics,
School of Mathematical and Natural Sciences,
University of Venda
Thohoyandou, Limpopo
South Africa

Supervisor: **Dr Caston Sigauke**

Co-Supervisor: **Dr Alphonce Bere**

May 2018

Declaration

I, Thakhani Ravele (student number 11630148), declare that the dissertation titled: "Medium term load forecasting in South Africa using generalized additive models with tensor product interactions" hereby submitted for qualification of Masters degree in Statistics at the University of Venda has not been submitted previously for any other degree at this university or any other institution. This dissertation is my own work and all significant references from the work of other authors have been duly acknowledged.

Signature (Student):..... Date:.....

Abstract

Forecasting of electricity peak demand levels is important for decision makers in Eskom. The overall objective of this study was to develop medium term load forecasting models which will help decision makers in Eskom for planning of the operations of the utility company. The frequency table of hourly daily demands was carried out and the results show that most peak loads occur at hours 19:00 and 20:00, over the period 2009 to 2013. The study used generalised additive models with and without tensor product interactions to forecast electricity demand at 19:00 and 20:00 including daily peak electricity demand. Least absolute shrinkage and selection operator (Lasso) and Lasso via hierarchical interactions were used for variable selection to increase the model interpretability by eliminating irrelevant variables that are not associated with the response variable, this way also overfitting is reduced. The parameters of the developed models were estimated using restricted maximum likelihood and penalized regression. The best models were selected based on smallest values of the Akaike information criterion (AIC), Bayesian information criterion (BIC) and Generalized cross validation (GCV) along with the highest Adjusted R^2. Forecasts from best models with and without tensor product interactions were evaluated using mean absolute percentage error (MAPE), mean absolute error (MAE) and root mean square error (RMSE). Operational forecasting was proposed to forecast the demand at hour 19:00 with unknown predictor variables. Empirical results from this study show that modelling hours individually during the peak period results in more accurate peak forecasts compared to forecasting daily peak electricity demand. The performance of the proposed models for hour 19:00 were compared and the generalized additive model with tensor product interactions was found to be the best fitting model.

Keywords: Generalized additive models, Lasso, Lasso via hierarchical interaction, medium term load forecasting, penalised regression, restricted maximum likelihood, tensor product interactions, time series.

Dedication

This work is heartily dedicated to my grandfather and grandmother (Mr. M.G. Mungoni Muhali and Bishop. T.G. Mungoni Muhali) for all the efforts, support and the difficulties that they have been through for the sake of my survival and success.

Acknowledgements

Firstly I would like to thank the almighty God for the gift of life, mercy and wisdom. My deepest gratitude to Dr. C. Sigauke, my supervisor, for his patience, inspiration, guidance through out this study and for introducing me to generalized additive models with tensor product interactions. My grateful thanks are also extended to Dr. A. Bere, my co-supervisor for his suggestions and guidance on this study. This project would not have been a success without the data that was provided by Eskom which is hereby highly acknowledged.

I am grateful to my parents and my sister for their patience, love and financial support. I appreciate them for being the pillars of my strength in encouraging me not to allow people's opinions become the focus of my meditation.

I wish to thank my classmates: Ms P. Mpfumali, Mrs T.P Nevhungoni and Mrs N. Nengovhela (Masters Statistics class of 2017&8) for encouraging me to work hard. The courageous words of Mr D. Mashavhela and Ms D. Netshivhumbe led to the whole idea of furthering my masters studies. I am grateful for their encouragement. The financial assistance from the CoE-MaSS and CSIR-DST bursaries of South Africa towards this study is hereby acknowledged. I sincerely thank the support of Mr L.W. Mulaudzi and Mr H. Nemavhulani for transporting me from home to school during dark days of this study. Without them this work would never have come into existence.

Table of contents

List of Figures

List of Tables

List of abbreviations

AIC	Akaikes Information Criteria
AIC_c	Akaikes Information Criteria corrected
ANNs	Artificial Neural Networks
BIC	Bayesian Information Criteria
CV	Cross Validation
DAH	Day after holiday
DBH	Day before holiday
DH	Day of holiday
DPED	Daily peak electricity demand
EF	Exponential family
GAM	Generalized Additive Model
GCV	Generalized cross validation
GLM	General Linear Model
LASSO	Least Absolute Shrinkage and Selection Operator
LL	Lower limit
LTLF	Long Term Load Forecasts
MAE	Mean Absolute Error
MAPE	Mean Absolute Percentage Error
MLE	Maximum Likelihood Estimation
MLR	Multiple Linear Regression
MSE	Mean Square Error
MTLF	Medium Term Load Forecasts
PDF	Probability Density Function
PIRLS	Penalized Iteratively Re-weighted Least Squares
PI	Prediction interval
PICP	Prediction interval coverage probability
PINAD	Prediction interval normalized average deviation
PINAW	Prediction interval normalized average width

PIW	Prediction interval width
Q-Q	Quantile-Quantile
R	Range
REML	Restricted Maximum Likelihood
RMSE	Root Mean Squared Error
SARMA	Seasonal autoregressive moving average
SARIMA	Seasonal autoregressive integrated moving average
STLF	Short Term Load Forecasts
UL	Upper limit

Chapter 1

Introduction

The future depends on decisions we take today. Decision makers in the energy sector are facing a serious crisis on planning for future world energy demand. Saving energy is an important task for energy planners and policy-makers. Consumption of energy increased worldwide for the past 10 years. Energy planners are facing challenges on how to maintain future economic prosperity and environmental security (Suganthi and Samuel, 2012). The environment we live in today requires energy resources to satisfy the world's future energy demands.

Nakata, (2004) modelled the change of energy systems by using the latest energy consumption data which influence our decision-making for future energy consumption. Electricity load forecasting faces rising challenges due to the advent of innovating technologies such as smart grids, electric cars and renewable energy production (Goude et al., 2014). Electricity load forecasting is important for production planning and trading on the electricity markets.

This study presents electricity load forecasting in South Africa. Electricity is the most used form of energy.

Forecasting is the process of predicting future events and conditions. Load forecasting is the prediction of future load at different time intervals. Load forecasting is used by energy planners to plan on how to maintain a balance between demand and supply of electricity. It has many applications such as energy purchasing and generation, load switching, contract evaluation, and infrastructure development. Load forecasting helps an electric utility management in planning the distribution of electricity. Load forecasting is also important for contract evaluations and evaluations of various sophisticated financial products on energy pricing offered by the market (Cavallaro, 2005; Singh et al., 2014; Mohandes, 2002).

The purpose of load forecasting is to give a prediction of hourly, daily, weekly and yearly demand or peak demand. The type of load forecasting can vary depending on the end result desired. It may be spatial if it mainly relates to the study of future patterns in a specific region, country or state. Otherwise it is temporal if concerned with forecasting on an hourly, daily, monthly or yearly basis. There is no best way of grouping load forecast categories. We can classify them according to their time horizons such as short term load forecasts (STLF) which are usually from one hour to one week, medium term load forecasts (MTLF) which are usually from a week to a year, and

long term load forecasts (LTLF) which are longer than a year (Feinberg and Genethliou, 2005; Hyndman and Fan, 2010; Hong and Fan, 2016; among others).

Electricity load forecasting has received a lot of attention from industrialists and academics in recent years (Goude et al., 2014). In this study we use the term "load forecasting" which refers to "electricity load forecasting", to avoid ambiguous and verbose presentation. The drivers of electricity demand are calendar variables which include the holiday effect and day of the week among others, weather data such as temperature, and possibly customer's classes. Many studies in the literature have focussed on STLF because of its implications on power system control, unit commitment, economic dispatch and electricity markets.

Important drivers of MTLF and LTLF are demographic factors such as population growth, weather variables such as temperature, changing technology, random usage of electricity by individuals etc. An overestimate of future load results in wastage of investment in construction of excess power facilities, while an underestimate of future load results in insufficient generation and unmet demand (Feinberg and Genethliou, 2005; Hyndman and Fan, 2010; Xie et al., 2016; Hong et al., 2016).

We use the term "technique" to represent a group of models that fall in

the same family, such as Multiple Linear Regression (MLR) models and Artificial Neural Networks (ANNs). Load forecasting techniques are categorised into: statistical techniques and artificial intelligence techniques. These techniques can be applied on load forecasting methods such as regression, neural networks, fuzzy logic, and expert systems. MTLF and LTLF use econometric and end-use methods in modeling. Methods employed for STLF are similar day approach, regression based models, time series, neural networks, statistical learning algorithms, fuzzy logic, and expert systems (Feinberg and Genethliou, 2005; Hong and Fan, 2016; Hyndman and Fan, 2010).

1.1 Problem Statement

South Africa is one of the developing countries and most industrialized on the African continent. South Africa is considered to be the highest electricity producer and consumer in Africa. It is estimated that more than half of the electricity generated in Africa comes from South Africa. The reserve margin estimate for 2008 is projected to be between 8-10%, which is far lower than the target set for the country, i.e. a minimum of 15% (Odhiambo, 2009). The South African government attributes the unprecedented decline in the electricity reserve margin during the recent years to the robust economic growth and the associated increase in demand for electricity (Odhiambo, 2010). Odhiambo (2009) found that there is a directly proportional relationship between electricity consumption and economic growth in South Africa. The modeling framework which will be developed in this study will

help decision makers in Eskom in medium term planning of the operations of the utility company.

1.2 Purpose of the study

1.2.1 Aim

The main aim of this study is to develop medium term load forecasting models which will help decision makers in Eskom in planning the operations of the utility company.

1.2.2 Objectives

The main objectives of this study are to:

- develop a hybrid medium term load forecasting model which combines the generalized additive model with tensor product interactions,
- estimate the parameters of the developed model using restricted maximum likelihood and penalized regression,
- evaluate the performance of the model,
- forecast hourly electricity demand during the peak period,
- evaluate the accuracy of the forecasts,
- suggest policy implications of this study.

1.3 Scope of the dissertation

This study is motivated by the work of Goude et al.(2014) in which they use generalized additive models (GAMs) in modeling the relationship between load and its driver variables. Similarly, the purpose of this study is to estimate the relationship between load demand and explanatory variables such as temperature, calendar effects, including lagged demand effects. Tensor product interactions will be used to analyze how the load forecast is influenced by interactions of explanatory variables.

The data which is going to be used in this study are obtained from Eskom. Eskom is South Africa's national public power utility company. The hourly historical load data is collected from January 2009 till December 2013. Generalized additive models with tensor product interactions will be used to estimate the relationship between load and the explanatory variables: temperature, calendar variables, etc. In addition, penalized cubic splines will be used to smooth the effects between the response variable with the predictor variables. Parameters of the developed GAMs will be estimated using restricted maximum likelihood estimation (RMLE) and penalized regression. The statistical analysis will be performed using R statistical packages which include among others, "gam" developed by Hastie and Hastie (2017), "mgcv" by Wood (2017).

1.4 Structure of the dissertation

In Chapter 2, we review existing literature on MTLF techniques as well as the summary of studies that use methods similar to the proposed one. Chapter 3 provides general theory of the statistical methods that will be used in this study. Chapter 4 provides data analysis while Chapter 5 presents the research conclusion.

Chapter 2

Literature review

2.1 Introduction

Research on medium term load forecasting(MTLF) for all sectors of the economy seems to be rare (Berk and Müller, 2016). A lot of research has been done on power system load forecasting. Most of the research focuses on STLF for predicting future hourly and daily loads, and LTLF which provides energy and peak demand forecasts for one or more years (Ghiassi et al., 2006; Xie et al., 2016; among others). In this chapter we give a review of some related literature. We provide an overview on MTLF, factors affecting it as well as the summaries of some studies that use the proposed methodology to forecast electricity demand in different countries.

2.2 Factors influencing medium term load forecasting

The MTLF takes into consideration customers demands, weather data, appliances used by customers and their characteristics including age of the appliences, the economic and demographic data and their forecasts, the appliance sales data, and other factors (Feinberg and Genethliou, 2005). Numerous factors which influence load demand considerably are economic, weather, time and seasonal variations, electricity prices, random disturbances, and other factors such as geographical location (urban or rural areas) etc. The most important weather variable is temperature which greatly influences MTLF. However, time and seasonal factors change the MTLF pattern (Singh et al., 2014).

The time factors include the time of the year, the day of the week, and the hour of the day. There are important differences in load demand between weekdays and weekends. The demand on different weekdays also can behave differently. During peak periods the supply, and demand fluctuate. Changes of weather conditions cause an increase in energy prices (Feinberg and Genethliou, 2005). MTLF also depends mainly on growth factors, such as main events, addition of new loads, demand patterns of large facilities, and maintenance requirements of large consumers. These types of forecasts are not concerned with hourly loads like short term forecasts, but rather pre-

dicts electric load for several months ahead.

Weather plays a huge role for prediction in MTLF. Researchers identified that the socioeconomic condition of some regions may rapidly change, and thus has impact on energy demands. For example, fast growing economies, migration or religious events such as pilgrimage periods. The MTLF often use economic indicators and electrical infrastructure measures in addition to historical load data and weather related variables to forecast future energy demands. However, the choice of weather variables(temperature measures, humidity factors, cooling degree days, duration of bright sunshine, wind speed, etc.) still depend on the regional weather characteristics (Ghiassi et al., 2006).

2.3 An overview of medium term load forecasting

MTLF is used for predicting future observations from 1 month up to 2 years. The forecasts provide useful information for power system planning and operations, and offer significant benefits for firms operating in a regulated or deregulated energy industry. MTLF can be used to optimize energy production and transmission, and improve overall system reliability. Moreover, it uses hourly loads for prediction of peak loads for several weeks ahead. These forecasts are used for planning the operations of electric utilities (Samuel et al., 2014; Ghiassi et al., 2006; Abu et al., 2011). Goude et al. (2014) propose

a model for medium term load forecasting, which is given by:

$$\begin{aligned} y_t = \sum_{j=1}^{7} m_j I_{DayType_{t=j}} + kI_{SpecialTariff_{t=1}} + g_1(\theta_t) + g_2(T_t) \\ + g_3(T_{t-1}) + g_4(T_{t-2}) + \sum_{j=1}^{11} O_j I_{OffSet_{t=j}} + h(toy_t) + \varepsilon_t \end{aligned} \tag{2.3.1}$$

where y_t denotes the demand of electricity at time t. In this formula, "DayType$_t$" is day classification for observation t: 1 for Sunday, 2 for Monday, 3 for Tuesday-Wednesday-Thursday, 4 for Friday, 5 for Saturday and 6 and 7 for Bank holidays. SpecialTarriff$_t$ is a factorial predictor, taking values 0 when there is special tariff at time t and 1 otherwise. θ_t denotes an exponential smoothing temperature of the real temperature T_t : $\theta_t = (10.99)T_t + 0.99\theta_{t-1}$, where T_{t-1} and T_{t-2} represent lag 1 and lag 2 temperature of one and two day(s) before. The categorical variable indicating the holidays and daylight saving time are denoted by $OffSet_t$. toy_t denotes date of the observation t within the year starting from 0 Jan the 1^{st} to 1 Dec 31^{st} and $h(toyt)$ corresponds to the variation of the yearly cycle of the load per instant of the day. However it is not easy to model MTLF since it is often driven by a lot of unknown covariates (Goude et al., 2014). Few models on medium term load forecasting have been proposed, to forecast electricity demand in different countries with the aim of finding models that will produce the best or accurate load estimates.

2.3.1 Generalized additive model (GAM)

Goude et al. (2014) use GAMs to model electricity demand for more than 2200 substations of the French distribution network, at both short and middle term horizons. The GAMs estimate the relationship between load and the explanatory variables: temperatures, calendar variables, etc. Goude et al. (2014) proposed the general model for GAM, which is given by;

$$y_t = f_1(x_{1t}) + f_2(x_{2t}) + ... + f_p(x_{pt}) + \xi_t \tag{2.3.2}$$

From equation 2.3.2, y_t is a univariate response variable, x_{pt} are the covariates that drive f_i, the smooth functions. The non-linear functions are supposed to be smooth, which means here that they can be relatively well estimated by penalized regression in a spline basis (Goude et al., 2014). GAMs are suitable for exploring the data set and visualizing the relationship between the dependent and independent variables (Liu, 2008). Pierrot and Goude (2011) use GAM to model effects that drove the French hourly load consumption and also compare the proposed model with the operational one. They estimate a GAM which includes the effects of different variables, and is given by;

$$\begin{aligned} L_t = {} & f_1(L_{t-24}) + f_2(L_{t-168}) + f_3(T_t) + f_4(mean(T_{t-24}, T_{t-48})) \\ & + f_5(cc) + f_6(posan) + C + \varepsilon_t \end{aligned} \tag{2.3.3}$$

where L_t, L_{t-24} and L_{t-168} are respectively the t instant load to be forecasted, the one-day-lagged load and the one-week-lagged load. T_t, T_{t-24} and T_{t-48} are the t instant temperature, the one-day-lagged temperature and the two-day-

lagged temperature. The cloud cover is denoted by *cc*, *posan* the position of the day through the year (from September to August), C is an intercept and ε_i the residual error.

Shadish et al. (2014) use the GAMs to examine whether the functional form trend exist between two variables. If the trend exists, then the GAM also determines its shape; i.e. whether it is linear or nonlinear. By fitting a GAM on weekly mean load demands, Cho et al. (2013) compose the whole movement, seasonality and climate change variables. GAMs are good in an interpretability, flexibility/automation, and regularization. GAMs provide a regularised and interpretable solution. GAMs strike a nice balance between the interpretable, yet biased, linear model, and the extremely flexible, black box learning algorithms.

2.3.2 Tensor product interactions

Tensor product interactions are a technique used to present smoothing functions of several interacting variables(variables binding). The purpose of this technique is to form a set of explanatory variable pairs that can be combined together to perform interactions, i.e. temperature and season can be combined to influence load demand simultaneously (Smolensky, 1990). Tensor products examine any interactions of variables that are used in a model (Gu, 1997).

2.4 Conclusions from literature

There are many studies where GAMs have been used in load forecasting. However, many of these studies focus on the use of semi parametric models, which have parametric and non-parametric components. The GAM is a generalized version of the semi parametric model. Few studies have been done on MTLF using GAM with tensor product interactions. In this study tensor product interactions will help us to examine the influence of several variables simultaneously. This study intends to make a contribution in the use of generalized additive models with tensor product interactions.

Chapter 3

Methodology

3.1 Introduction

This chapter discusses the methods and techniques used in analysing the data. The generalized additive model (GAM) and tensor product interactions will be discussed including the application of generalized additive model with tensor product interactions. Maximum likelihood estimation and the use of residuals in checking violations of assumptions will be discussed. We will also discuss how the models are evaluated using accuracy measures.

3.2 Generalized Additive Model

Generalized additive models (GAMs) are the models in which the relationships between independent and dependent variables are additive. GAMs were initially introduced by Hastie and Tibshirani (1990). Generalized linear models (GLMs) assume a functional relationship between the response and the predictor variables. GAMs do not assume a functional relationship

apriori and as such allow for more flexibility in the modelling. GAMs include nonlinear form of predictors and use smooth functions of predictor variables in modelling which can be applied in different forms. A GAM is given as:

$$g(\mu_t) = \mathbf{A}\boldsymbol{\theta} + \sum_{j=1}^{p} f_j(x_{jt}) + \xi_t, \quad \mathrm{Y}_t \sim \mathrm{EF}(\mu_t, \phi), \quad t = 1, ..., n \tag{3.2.1}$$

where Y_t denotes independent univariate response variables from Gamma exponential family distribution with mean μ_t and scale parameter ϕ, smooth monotonic link function is denoted by g, $\mathbf{A}$ is a design matrix of order $n \times k$, $\boldsymbol{\theta}$ denotes an unknown parameter vector of order $k \times 1$, f_j is an unknown smooth function of the predictor variable x_j that may be vector valued, and ξ_t is an independent identical distribution random error $k \times 1$ (Chouldechova and Hastie, 2015; Hastie and Tibshirani, 1990; Pya and Wood, 2016). The non-linear (non-parametric) functions are assumed to be smooth. It can be relatively well estimated by penalised regression in a spline basis. Each function is given by (Goude et al., 2014):

$$f_j(x_{jt}) = \sum_{t=0}^{k_j} \beta_{j,t} b_t^j(x_{jt})$$

where k_j denote the dimension of the spline basis to model the effect $f_j(x_{jt})$ and $b_t^j(x_{jt})$ is the corresponding spline functions, i.e. B-splines or cubic regression splines and $\beta_{j,t}$ are coefficients to be estimated.

3.3 Tensor product interactions

According to Laurinec (2017) interactions are very important part of the regression model for double seasonal time series. They can be applied to

a GAM using: product of explanatory variables $x_1 \times x_2 \times x_3 \times ... \times x_n$, smoothed function or one variable $f_1(x_1) \times x_2$ and smoothed function for both variables $f_1(x_1) \times f_1(x_2)$. The interactions of two independent variables using tensor product is given by $f_1(x_1) \bigotimes f_2(x_2)$ where $\bigotimes$ is a kronecker product (Laurinec, 2017). Tensor product interactions produce smooth functions of several variables (Wood, 2006). Wood (2006) developed smooth functions of two covariates, x_i and x_k by assuming low rank bases. The smooth function of each covariate is denoted by f_i and f_k. The smooth functions are given respectively by:

$$f_i(x_i) = \sum_{i=1}^{I} \alpha_i a_i(x_{ti}) \tag{3.3.1}$$

$$f_k(x_k) = \sum_{k=1}^{K} \beta_k b_k(x_{tk}) \tag{3.3.2}$$

where α_i and β_k are parameters, and $a_i(x)$ and $b_k(x)$ are known basis functions. A general smooth function combining two covariates is given by:

$$f_{ik}(x_i, x_k) = \sum_{i=1}^{I} \sum_{k=1}^{K} \beta_{ik} a_i(x_{ti}) b_k(x_{tk}) \tag{3.3.3}$$

where $(\beta_{ik})_{i,k=1}^{I,K}$ can be any predictor variables but i predictor variable cannot be the same as k predictor variable simultaneously $(i \neq k)$. In this study we consider pairwise interactions only, however this can easily be extended to higher order interactions. We propose interactions because they will help us to model the influence of several predictor variables simultaneously.

3.4 Generalized additive model with tensor product interactions

Let Y_t denote electricity demand at time t, then the GAM with tensor product interactions is given as:

$$Y_t = \mathbf{A}\theta + \sum_{j=1}^{p} f_j(x_{jt}) + \sum_{i=1}^{I}\sum_{k=1}^{K} \beta_{ik} a_i(x_{ti}) b_k(x_{tk}) + \xi_t \tag{3.4.1}$$

where $t = 1, ..., n$, $j \neq k$, ξ_t is an error term which is assumed to follow the seasonal autoregressive moving average (SARMA) model and its expression is given as:

$$\phi(B)\Phi(B^s)\xi_t = \theta(B)\Theta(B^s)a_t \tag{3.4.2}$$

where a_t is white noise. Now from 3.4.1

$$\xi_t = Y_t - [\mathbf{A}\theta + \sum_{j=1}^{p} f_j(x_{jt}) + \sum_{i=1}^{I}\sum_{k=1}^{K} \beta_{ik} a_i(x_{ti}) b_k(x_{tk})]$$

and we substitute ξ_t into the SARMA model, to get

$$\phi(B)\Phi(B^s)\{Y_t - [\mathbf{A}\theta + \sum_{j=1}^{p} f_j(x_{jt}) + \sum_{i=1}^{I}\sum_{k=1}^{K} \beta_{ik} a_i(x_{ti}) b_k(x_{tk})]\} = \theta(B)\Theta(B^s)a_t \tag{3.4.3}$$

3.5 Variable selection

Variable selection is a process of selecting best predictors or set of predictors. Variable selection helps in removing the unnecessary predictors and avoid overfitting. There are many methods for variable selection, however in this study we use Lasso (least absolute shrinkage and selection operator) and

Lasso via hierarchical interactions. Lasso was first introduced by Tibshirani (1996). Lasso is a method that performs both model/variable selection and estimation in order to improve the prediction accuracy (Bien et al., 2013; Lim and Hastie, 2015).

Lasso variable selection sets a constraint on the sum of the absolute values of the model parameters, the sum has to be less than a fixed value (upper bound). The method apply a shrinking (regularization) process where it penalises the coefficients of the regression variables shrinking some of them to zero. The non-zero coefficient after the shrinking process are selected to be part of the model. The goal of Lasso variable selection is to minimise the prediction error (Fonti and Belitser, 2017). We consider pairwise hierarchical interactions only, however this can easily be extended to higher order interactions. We assume a regression model with response variable Y_t and predictors X_1,...,X_p with pairwise interactions between these predictors. The hierarchical model is given as (Bien et al., 2013):

$$Y = \beta_0 + \sum_j \beta_j X_j + \frac{1}{2} \sum_{j \neq k} \Theta_{jk} X_j X_k + \xi \tag{3.5.1}$$

where $\xi \sim N(0, \sigma^2)$, β_0 and β_j denote parameters to be estimated. Θ_{jk} is a symmetric matrix of interactions. Irrespective of whether the predictors are continuous or discrete, the additive part is treated as the "main effect" terms and the quadratic part is treated as the "interactions" terms. The main idea is to estimate $\beta \epsilon R^p$ and $\Theta \epsilon R^{p*p}$, where $\Theta \epsilon \Theta^T$ and Θ_{jj} =0. In this study Θ

is used for interactions instead of using vector of length $\frac{1}{2}p(p-1)$ and Θ_{jj} =0 throughout this study to simplify notations. There are two categories of hierarchical restrictions, which are strong and weak hierarchy (Bien et al., 2013).

$$Strong\ hierarchy: \hat{\Theta}_{jk} \neq 0 \Rightarrow \beta_j \neq 0\ and\ \hat{\Theta} \neq 0$$

$$Weak\ hierarchy: \hat{\Theta}_{jk} \neq 0 \Rightarrow \beta_j \neq 0\ or\ \hat{\Theta} \neq 0$$

The major advantage of lasso via hierarchical interactions is that it produces simpler and more interpretable models that involve only a subset of the predictors (Bien et al., 2013; Lim and Hastie, 2015). Once the variables have been selected the next step is to estimate the parameters.

3.6 Parameter estimation

Parameter estimation is when we estimate the parameters of the model using the given data. Parameter estimation has a huge role in accuracy description and its purpose is to evaluate the parameters obtained from any scenario. There are many methods for parameter estimation, however we will discuss four appropriate methods in this study which are: maximum likelihood estimation (MLE), restricted maximum likelihood (REML), penalized iteratively re-weighted least squares (PIRLS) and Bayesian estimation. MLE estimate the value of one or more parameters for a certain statistic which makes the unknown likelihood distribution a maximum and parameters are assumed to be fixed but unknown. REML is a type of MLE in which the estima-

tion of parameters do not focus on a maximum likelihood, however it uses a likelihood function calculated from a transformed set of data. Iteratively re-weighted least squares use MLE of GLM and a robust regression to identify the effect of outlier on a data set. Bayesian methods estimate the posterior density denoted by $p(\theta|X)$. In this method parameters are assumed to be random variables with some known prior distribution, and uncertainty about parameters are represented by a probability density function (pdf). In this study the restricted maximum likelihood and penalized regression methods are going to be used to estimate parameters because REMLs are best linear unbiased estimates (BLUE). REML is also used because it provides consistent estimates of parameters and several popular statistical software packages provide excellent algorithms for REML for most commonly used distributions. The principle of MLE starts with a sample of random variables. Suppose $X_1, X_2, X_3, ..., X_n$ have joint density function denoted by:

$$f_\theta(x_1, x_2, ..., x_n) = f(x_1, x_2, ..., x_n|\theta) \tag{3.6.1}$$

given observed values $X_1 = x_1, X_2 = x_2, ..., X_n = x_n$. The MLE of θ is given from Ross (2014) by:

$$L(x_i, \theta) = \prod_{i=1}^{n} f(x_i, \theta) \tag{3.6.2}$$

where is θ the value that maximises the $L(\theta)$. The logarithm is an increasing function which can also maximise the log likelihood, which is given by:

$$\ell(x_i, \theta) = \sum_{i=1}^{n} \log(f(x_i|\theta)) \tag{3.6.3}$$

A classical way to estimate the smooth effects is penalized regression, more precisely ridge regression, where we minimize the following criteria (Goude et al,. 2014):

$$\sum_{t=1}^{n}(y_t - \sum_{j=1}^{k} f_j(x_t))^2 + \sum_{j=1}^{k} \lambda_j \int \|f_j''(x)\|^2 dx \tag{3.6.4}$$

where the penalty parameter $\bigwedge = (\lambda_1 ... \lambda_k)$ which controls the degree of smoothness of each effect. The higher λ_j the smoother f_j which has to be optimized. $\mathbf{B}$ is the matrix formed by concatenation of the b_t^j. The minimization is given by:

$$min_{\beta,\lambda} \|\|Y - \mathbf{B}\|\|^2 + \sum_{j=1}^{k} \lambda_j \beta^T S_j \beta \tag{3.6.5}$$

where β denote the vector of the unknown regression parameters. S_j is a smoothing matrix depending on the spline basis.

3.7 Model selection

Model selection is a process of using the data to choose the best model from the list of candidate models. The best model is selected using information criteria (Wasserman, 2000). Information criteria has been established to summarize data to verify model and reduce the information needed to explain data and model. The useful information criteria is Akaikes Information Criterion (AIC). This technique developed by Akaike (1973), aim to choose the model that minimises the negative likelihood penalised by the number of parameters. The idea of AIC is to find the best approximating model to the

unknown true data generating process and its applications (Akaike, 1973). The AIC equation is given by:

$$\text{AIC} = 2k - 2\log(\ell(\theta)) \tag{3.7.1}$$

AIC with a correction for finite sample sizes is an akaike information criterion corrected (AIC_c). AIC_c is given by:

$$\text{AIC}_c = \text{AIC} + \frac{2k(k+1)}{n-k-1} \tag{3.7.2}$$

Another widely used information criteria is the Bayesian Information Criterion (BIC). Schwarz (1978) developed BIC to find the most probable model given the data. BIC is given by:

$$\text{BIC} = k\log(n) - 2\log(\ell(\theta)) \tag{3.7.3}$$

where, k is the model degrees of freedom, n is the number of observations and $\ell(\theta)$ is denotes likelihood function of parameter (θ). A simple heuristic approach for model selection is Cross Validation (CV). CV is used for selection of prediction models and it divides sample into two subsets, such as a training (calibration) subset for modeling fitting and a test (validation) subset for model evaluation. This method is efficient when the sample is large (Liu, 2008). CV formula is given by (James et al, 2013):

$$CV_n = \frac{1}{n}\sum_{t=1}^{n}(y_t - \hat{y}_t)^2 = \frac{1}{n}\sum_{t=1}^{n} MSE_t \tag{3.7.4}$$

where MSE_t denotes the mean square error, which provides an approximately unbiased estimate for the test error. CV is extended to generalized

cross validation (GCV) and adjusted R^2 are the types of CV creterion. The GCV criterion measure the goodness of fit which consider the residual error and the model complexity. GCV criterion is given by (Craven and Wahba, 1979):

$$GCV(M) = \frac{\frac{1}{n}\sum_{t=1}^{n}[y_t - \hat{f_M}(x_t)]^2}{[1 - \frac{G(M)}{n}]^2} \tag{3.7.5}$$

where n denotes the sample size, and C(M) is the cost-penalty measure of a model containing M basis functions. The numerator determines the lack of fit on the M basis function model $\hat{f_M}(x_t)$ and the denominator represents the penalty for the model complexity C(M). The best model is one with the lowest GCV criterion value. In adjusted R^2, the best model is the one with the highest value or highest percentage explained by the model. Adjusted R^2 is given by (Hyndman and Athanasopoulos, 2014):

$$Adjusted R^2 = 1 - \frac{(1 - R^2)(n - 1)}{n - k - 1} \tag{3.7.6}$$

3.8 Model diagnostics

Model diagnostics is a process of checking whether the assumptions of the residuals are violated or not. Examination of whether the residuals are normally distributed and independent will be carried out. Violations of the assumptions may result in an unstable model leading to wrong conclusions. A residual is the difference between the response variable and the estimated values of a response variable. The residuals definition is given by:

$$\xi_t = y_t - \hat{y}_t \tag{3.8.1}$$

where, y_t is the actual values of a response variable and $\hat{y}_t$ is estimated values of a response variable (Cook and Weisberg, 1982).

3.9 Evaluation of forecasts

3.9.1 Accuracy measures

Accuracy measures which are normally used to evaluate the forecasting accuracy of a prediction model are mean absolute error (MAE), mean absolute percentage error (MAPE) and root mean squared error (RMSE). MAE uses the same units as the data, and is usually similar in magnitude to, but slightly smaller than, the root mean squared error. Researchers indicate that the MAE is a simpler statistic to understand than the RMSE and MAPE. MAPE expresses the accuracy as a percentage and one of its drawbacks is that the measure is not defined for actual values equal to zero. RMSE is a measure of the square root of the average of squared errors. For large errors the effect is large hence the method is sensitive to outliers. MAE and RMSE are commonly used in model evaluation studies (Chai and Draxler, 2014). The equations of the above measures are given by:

$$\text{MAE} = \frac{1}{m}\sum_{t=1}^{m}|\xi_t| \tag{3.9.1}$$

$$\text{MAPE} = \frac{100}{m}\sum_{t=1}^{m}|\frac{\xi_t}{\hat{y}_t}| \tag{3.9.2}$$

$$\text{RMSE} = \sqrt{\frac{1}{m}\sum_{t=1}^{m}\xi_t^2} \tag{3.9.3}$$

where, m is the number of observations in the test data set, $\hat{y}_t$ is the estimated values of a response variable and ξ_t is the residual of the t^{th} observation, given as $\xi_t = y_t - \hat{y}_t$

3.9.2 Quantile regression averaging

Probability load forecasting yield electric load forecasting in the form of intervals, scenarios, density function or probability (Liu et al, 2017). In this study quantile regression averaging (QRA) is used to determine the prediction intervals for the forecasting models. The forecasts of the predicted models are treated as independent variable of actual demand. The QRA is given by:

$$y_{t,\tau(QRA)} = \beta_0 + \sum_{i=1}^{m} \beta_i \hat{y}_{ti} + \xi_t \tag{3.9.4}$$

where $\hat{y}_{ti}$ denotes load forecasts at time t.

3.9.3 Prediction interval

Prediction interval (PI) is a usefully tool for uncertainty modeling by its nature. PI composed of lower and upper bounds that cover the future unknown target value with a certain probability $(1-\alpha)\%$ called the confidence level. PIs are more suitable and producing useful information than point forecasts for decision makers (Quan et al, 2014; Shen et al, 2018). Prediction interval width (PIW) is defined by:

$$PIW_t = UL_t - LL_t \tag{3.9.5}$$

where UL_t and LL_t is the upper and lower limit. Too large width is not useful for making decision in the electrical power system. The typical indicators that evaluate the performance of PIs of the models are prediction interval coverage probability (PICP) and prediction interval normalized average width(PINAW). PICP examine the reliability of the constructed PIs and is the fundamental feature of PIs. PICP is useful measures the probability of targets lie in the constructed PIs. PICP is calculated as follows (Quan et al, 2014; Shen et al, 2018):

$$PICP_t = \frac{1}{m}\sum_{t=1}^{m} I_t \tag{3.9.6}$$

where m is the number of samples and

$$I_t = \begin{cases} 1 & \text{if} \quad y_t \epsilon [LL_t; UL_t] \\ 0 & \text{if} \quad otherwise \end{cases}$$

PICP is valid, if PICP is greater or equal to the nominal confidence level of PIs (PICP$\geq 1-\alpha$). PICP=100% if the target values are within the PIs. Prediction interval normalized average width (PINAW) is the significant evaluation index of the PIs. PINAW is the higher quality of PIs, which evaluate overall width of PIs. PINAW is given by (Quan et al, 2014; Shen et al, 2018):

$$PINAW = \frac{1}{mR}\sum_{t=1}^{m}(UL_t - LL_t) \tag{3.9.7}$$

where R is the range of the underlying targets. A higher PICP implies that more prediction points lie in the constructed PIs, which may lead to a smaller

prediction interval normalized average deviation (PINAD). PINAD state the deviation degree from the real values to the PIs. PINAD is given by (Sun et al, 2017):

$$PINAD = \frac{1}{mR}\sum_{t=1}^{m} d_t \tag{3.9.8}$$

where d_t denotes the deviation degree and is given by:

$$d_t = \begin{cases} LL_t - y_t & \text{if} \quad y_t < LL_t \\ 0 & \text{if} \quad LL_t \leq y_t \leq UL_t \\ y_t - UL_t & \text{if} \quad y_t > UL_t \end{cases}$$

3.9.4 Forecast error distribution

Forecast error distribution investigate the effect of descriptive (summary) statics on distribution (Barrow and Kourentzes, 2016). In this study forecast error distribution is used with reference for over and under prediction.

3.9.5 Percentage improvement

Percentage improvement define the improvement of best model relative to other models. Percentage improvement between best model with the other models is given by (Abuella and Chowdhury, 2017):

$$Improvement(\%) = (1 - \frac{MAE_{best\ model}}{MAE_{other\ model}}) * 100\% \tag{3.9.9}$$

where $\text{MAE}_{best\ model}$ is mean absolute error for best model and $\text{MAE}_{other\ model}$ is mean absolute error for other model.

Chapter 4

Data Analysis and discussion

4.1 Introduction

This chapter presents the data analysis using the methods and models discussed in Chapter 3. The statistical package used in the analysis of the data is RStudio. After this introduction follows a Section 4.2 that gives an overview of the data, then Sections 4.3, 4.4 and 4.5 which give the modelling of demand at hours 19:00 and 20:00, and DPED respectively. Section 4.6 compare forecasts of the proposed models. Section 4.7 presents operational forecasts of hour 19:00 and Section 4.8 compare the performance of proposed models. Lastly Section 4.9 presents some concluding remarks.

4.2 Data

4.2.1 Data sources

The study uses secondary data, obtained from Eskom, South Africa's power utility company. Eskom controls the generation, transmission and distri-

bution of electricity to the South African sectors. Approximately 95% of electricity used in South Africa is produced by Eskom. It has been estimated that more than half of the electricity generated in Africa comes from South Africa (Odhiambo, 2009). Aggregated hourly load data for the national system in South Africa for the period 01 January 2009 to 31 December 2013 is used in this study. The aggregated hourly load data is obtained from various sectors of South Africa economy, i.e. residential, industrial and commercial sectors.

4.2.2 Data preparation

The variables used in this study are:

- $y_{t,h}$ which represents the electricity demand on day t, t = 1,2,...,n at hour h, h = 1,2,...,24 and in this study we only consider h=19 and 20,
- DPED$_t$ is the daily peak electricity demand at time t,
- minED$_t$ and AveED$_t$ are the minimum and average electricity demand at time t,
- AminT$_t$, AveT$_t$ and AmaxT$_t$ are the minimum, average and maximum daily temperatures at time t,
- DayType is a factorial variable with seven levels that correspond to different days of the week, namely: 1 for Monday, 2 for Tuesday, 3 for Wednesday, 4 for Thursday, 5 for Friday, 6 for Saturday and 7 for Sunday,

- DBH_t, DH_t and DAH_t are day before holiday, day of holiday and day after holiday respectively,
- Month is the months of the year coded as 1 for January,..., 12 for December,
- trend is the trend variable.

The data is split into two sets: a training set, from 01 Jan 2009 to 31 Dec 2011, and a testing set, from 01 Jan 2012 to 31 Dec 2013.

4.2.3 Exploratory data analysis

The frequency analysis of different hours of the day using the historic data of the period 2009 to 2013 are shown in Figure 4.1 and Table 4.1. From Figure 4.1 most of the daily peak electricity demand occurs on hours 19:00 and 20:00 respectively. To avoid computational burden, the study is going to consider these two hours, 19:00 and 20:00.

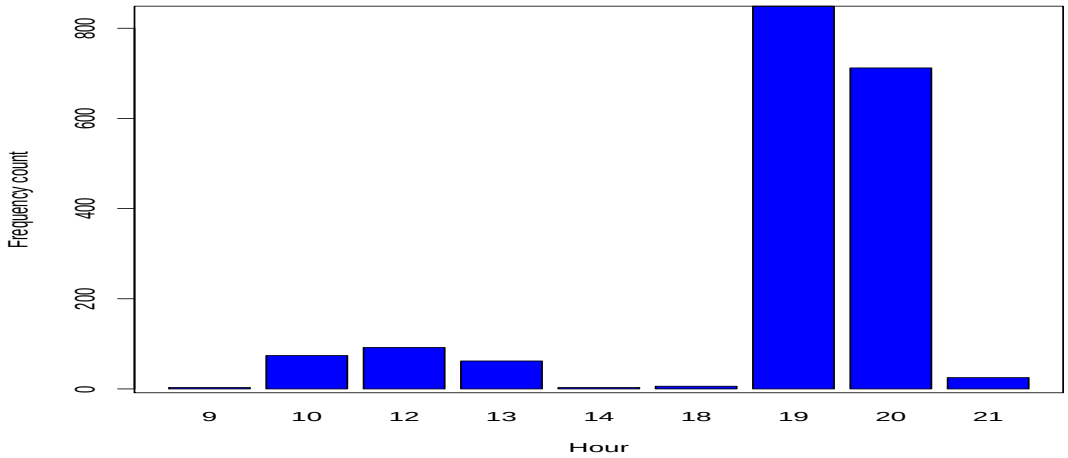

Figure 4.1: Bar chart of the daily electricity peak demand occurrences for the period 01 Jan 2009 to 31 Dec 2013.

Table 4.1: Frequency of occurrences of DPED in different hours of the day.

DPED Hour	9	10	12	13	14	18	19	20	21	Total
Frequency	3	74	92	62	3	6	849	712	25	1826
Percentage	0.16	4.05	5.04	3.40	0.16	0.33	46.50	38.99	1.37	100

It is presented in Figure 4.1 and Table 4.1 that the daily peaks of electricity demand in South Africa occur at hours 10:00, 12:00, 13:00, 19:00 and 20:00. Table 4.2 gives a summary of the descriptive statistics of five peak hourly electricity demand of the day. Since the values of skewness of hours 12:00 and 19:00 are between -0.5 and 0, the electricity demand of both hours is approximately symmetric. However the skewness value of hours 10:00, 13:00 and 20:00 are between -1 and -0.5, showing that the electricity demand of each hour is moderately skewed. The kurtosis values of all hours are greater

than 0 respectively, which means the electricity demand is a leptokurtic.

Table 4.2: Descriptive statistics of electricity demand in five peak hours of the day.

Hour	Min	Q_1	Q_2	Mean	Q_3	Max	Skew	Kurtosis
10	21866	28844	29923	29761	30820	34970	-0.583	1.267
12	22599	28801	29946	29798	30983	34789	-0.486	0.735
13	22281	28499	29615	29456	30639	34380	-0.576	0.816
19	20372	29219	31036	30927	32819	36970	-0.362	0.016
20	21948	29732	31301	31044	32443	36073	-0.625	0.844

Table 4.1 depicted that the hours 19:00 and 20:00 are the two hours in a day experiencing high electricity usage by consumers. Time series plots of the daily peak electricity demand at peak hours 19:00 and 20:00 together with their respective densities for the sampling period 01 Jan 2009 to 31 Dec 2013 are given in Figure 4.2. The electricity demand at both hours, 19:00 and 20:00 is steadily decreasing as from 2011. The possible reasons for decrease in electricity demand its due to integration of renewable energies, some big companies such as Sasol are using their own self generated electricity, etc.

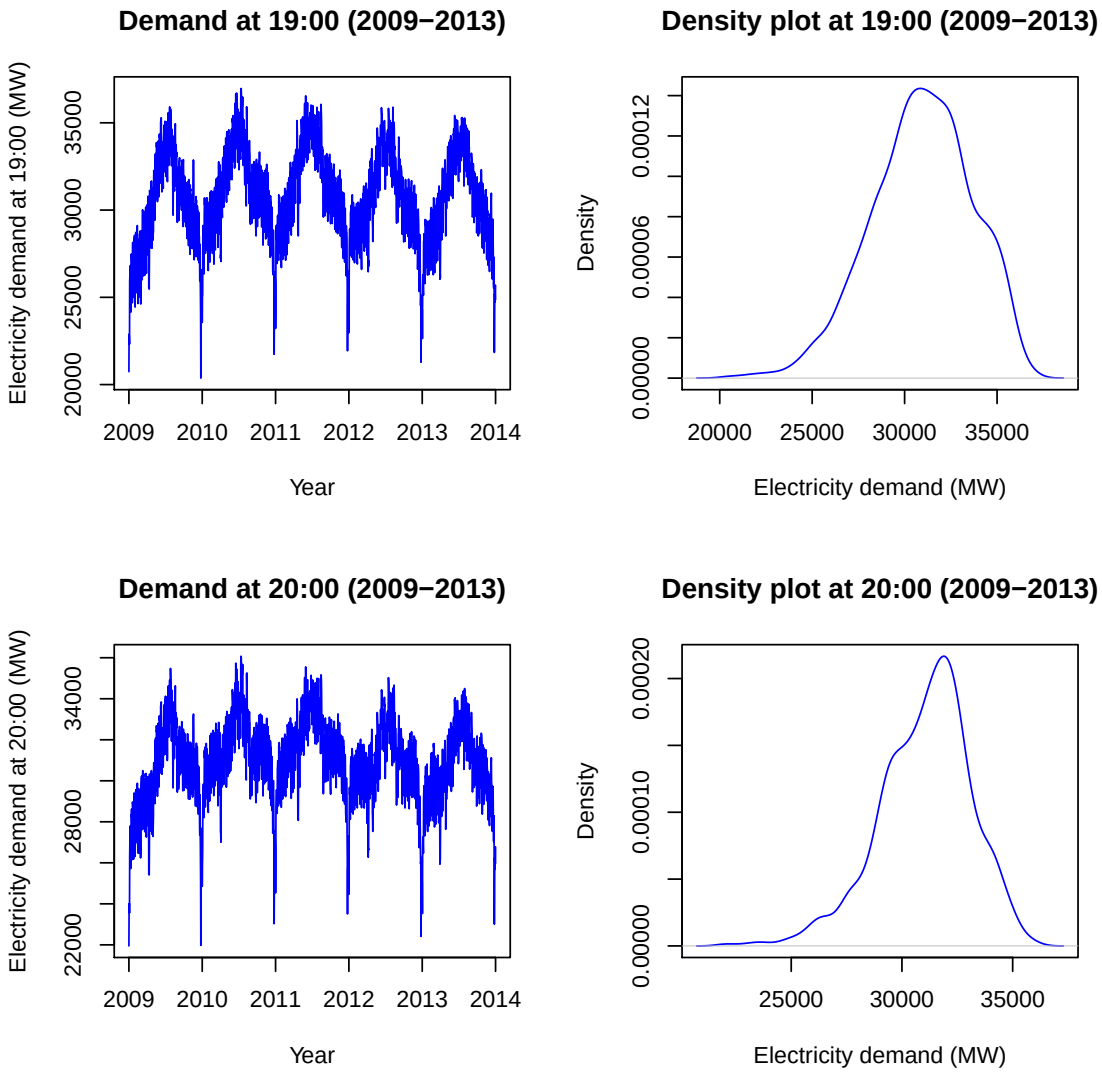

Figure 4.2: Electricity demand time series and density plots at hours 19:00 and 20:00.

4.3 Generalized additive model for 19:00

4.3.1 Introduction

The different features of the models which include variable description, variable selection and model building are discussed in this section.

4.3.2 Variable and model selection

A simple generalized additive model (GAM) which include all the predictor variables is considered before any variable selection using Lasso and Lasso via hierarchical interactions. The first GAM with main effects for 19:00 is given by:

$$\begin{aligned} y_{t,19} = \beta_0 + s_{1(19)}(Lag_1) + s_{2(19)}(Lag_2) + s_{3(19)}(\text{DPED}) + s_{4(19)}(\text{minED}) + \\ s_{5(19)}(\text{AveED}) + s_{6(19)}(\text{AminT}) + s_{7(19)}(\text{AveT}) + s_{8(19)}(\text{AmaxT}) + \\ s_{9(19)}(\text{DBH}) + s_{10(19)}(\text{DH}) + s_{11(19)}(\text{DAH}) + s_{12(19)}(\text{Month}) + \\ s_{13(19)}(\text{DayType}) + s_{14(19)}(\text{trend}) + \xi_{t(19)} \end{aligned} \tag{4.3.1}$$

where $y_{t,19}$, represents the electricity demand on day t at hour 19:00. β_0 and $s_{1(19)}, ..., s_{14(19)}$ are parameters to be estimated.

Variable selection using Lasso for 19:00

Model (4.3.1) is the reference model for Lasso variable selection including Lag_1 and Lag_2. Table 4.3 shows variables selection using Lasso for hour 19:00. From Table 4.3 the variables with zero coefficients are not selected.

Table 4.3: Variable selected using Lasso for 19:00 hour demand.

(Intercept)	-3.098335e+03
Lag_1	0
Lag_2	0
DPED	1.176217e+00
minED	-1.098256e-01
AveED	0
AminT	0
AveT	-2.276270e+01
AmaxT	-5.427992e+00
DBH	1.267646e+02
DH	0
DAH	0
Month	0
DayType	0
trend	-6.424805e-02

Variables selected are DPED, minED, AveT, AmaxT, DBH and trend. GAM equation after Lasso variable selection is given by:

$$y_{t,19} = s_{1(19)}(\text{DPED}) + s_{2(19)}(\text{minED}) + s_{3(19)}(\text{AveT}) + s_{4(19)}(\text{AmaxT}) + s_{5(19)}(\text{DBH}) + s_{6(19)}(\text{trend}) + \xi_{t(19)} \tag{4.3.2}$$

Variable selection using Lasso via hierarchical interactions for 19:00

In this section we are motivated by the study of Bien et al.(2013), where they used two-way interactions which include main effect terms and interactions. Table 4.4 shows variable selection using Lasso via hierarchical interactions. The variables that has zero at the cell in which they are interacting with other variables, it means the variables are not interacting. In other words there is no interactions between the two specific variables.

Table 4.4: Variable selected using Lasso via hierarchical interactions for demand at hour 19:00.

	Main effect	Lag_1	Lag_2	DPED	AmaxT	Tight?
Lag_1	59.961	-32.4488	0	-10.1188	17.3934	*
Lag_2	75.9533	0	0	-51.5014	3.261	
DPED	2275.8755	-10.1188	-51.5014	0	0	
AmaxT	-28.8817	17.3934	3.261	0	0	

From Table 4.4, interacting variables are Lag_1 interacts with DPED, Lag_1 interact with AmaxT, Lag_2 interact with DPED and Lag_2 interact with AmaxT. The GAM after variable selection using Lasso via hierarchical interactions for 19:00 is given by:

$$\begin{aligned} y_{t,19} = \beta_0 + s_{1(19)}(Lag_1) + s_{2(19)}(Lag_2) + s_{3(19)}(\text{DPED}) + s_{4(19)}(\text{minED}) + \\ s_{5(19)}(\text{AveED}) + s_{6(19)}(\text{AminT}) + s_{7(19)}(\text{AveT}) + s_{8(19)}(\text{AmaxT}) + \\ s_{9(19)}(\text{DBH}) + s_{10(19)}(\text{DH}) + s_{11(19)}(\text{DAH}) + s_{12(19)}(\text{Month}) + s_{13(19)}(\text{DayType}) + \\ s_{14(19)}(\text{trend}) + s_{15(19)}(\text{Lag}_1,\text{DPED}) + s_{16(19)}(\text{Lag}_1,\text{AmaxT}) + \\ s_{17(19)}(\text{Lag}_2,\text{DPED}) + s_{18(19)}(\text{Lag}_2,\text{AmaxT}) + \xi_{t(19)} \end{aligned} \tag{4.3.3}$$

4.3.3 Model diagnostics

The Ljung Box test is used to check for autocorrelation in the fitted GAM. When there is no autocorrelation the p-value of Ljung Box test must be greater than 0.05. If p-value is less than 0.05, autocorrelation is reduced until p-value is greater than 0.05. GAM (4.3.2), with main effects is used for checking autocorrelation.

After testing for autocorrelation, the p-value of residuals is 0.00 (p=0.00). Since the p-value is less than 0.05, we conclude that there is autocorrelations in the residuals. Furthermore we reduce the autocorrelation, to make p-value greater than 0.05. In order to reduce the autocorrelation, SARIMA(5,0,2)(2,0,3)$_7$ is used and the autocorrelation is reduced until getting p-value equal to 0.33 (p=0.33). Since p-value is greater than 0.05, we conclude that there is no autocorrelation in the residuals at 5% level of significance.

Secondly the GAM (4.3.3), with the main effects and interactions is considered to check the residuals autocorrelation. After testing for autocorrelation, the p-value of residuals is 0.00 (p=0.00). Since the p-value is less than 0.05, we conclude that there is autocorrelations in the residuals. Furthermore we reduce the autocorrelation, to make p-value greater than 0.05. In order to reduce the autocorrelation, SARIMA(5,0,2)(2,0,3)$_7$ is used and the autocor-

relation is reduced until getting p-value equal to 0.49 (p=0.49). Since p-value is greater than 0.05, we conclude that there is no autocorrelation in the residuals at 5% level of significance. SARIMA$(5,0,2)(2,0,3)_7$ is given by equation (4.3.4):

$$\begin{aligned}\xi_t = &\sum_{i=1}^{5}\theta_i\xi_{t-i} - \Phi_1\sum_{i=1}^{5}\theta_i\xi_{t-i-7} + \Phi_2\sum_{i=1}^{5}\theta_i\xi_{t-i-14} - \Phi_1\xi_{t-1} - \Phi_2\xi_{t-2} - \\ &\sum_{i=1}^{2}\theta_i a_{t-i} + \Theta_1\sum_{i=1}^{2}\theta_i a_{t-i-7} + \Theta_2\sum_{i=1}^{2}\theta_i a_{t-i-14} + \Theta_1\sum_{i=1}^{2}\theta_i a_{t-i-21} - \\ &\sum_{i=1}^{3}\Theta_i a_{t-3i} + a_t\end{aligned} \tag{4.3.4}$$

Time series display of residuals for GAMs (4.3.2) and (4.3.3) before and after correcting autocorrelation are given in Figure 4.3 and 4.4 respectively.

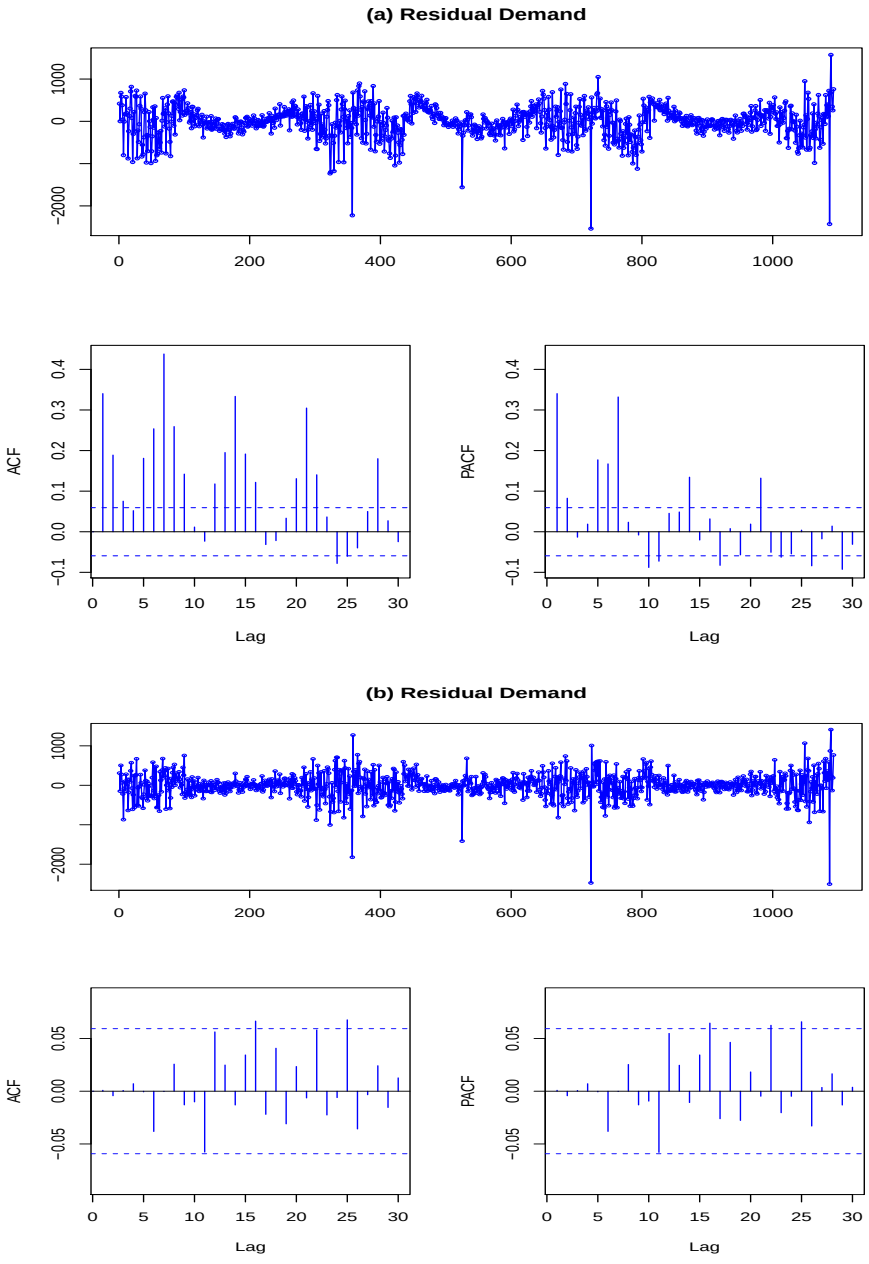

Figure 4.3: **Top**: Time series display for the residuals before reducing autocorrelation for model (4.3.2). **Bottom**: Time series display for the residuals after reducing autocorrelation for model (4.3.2).

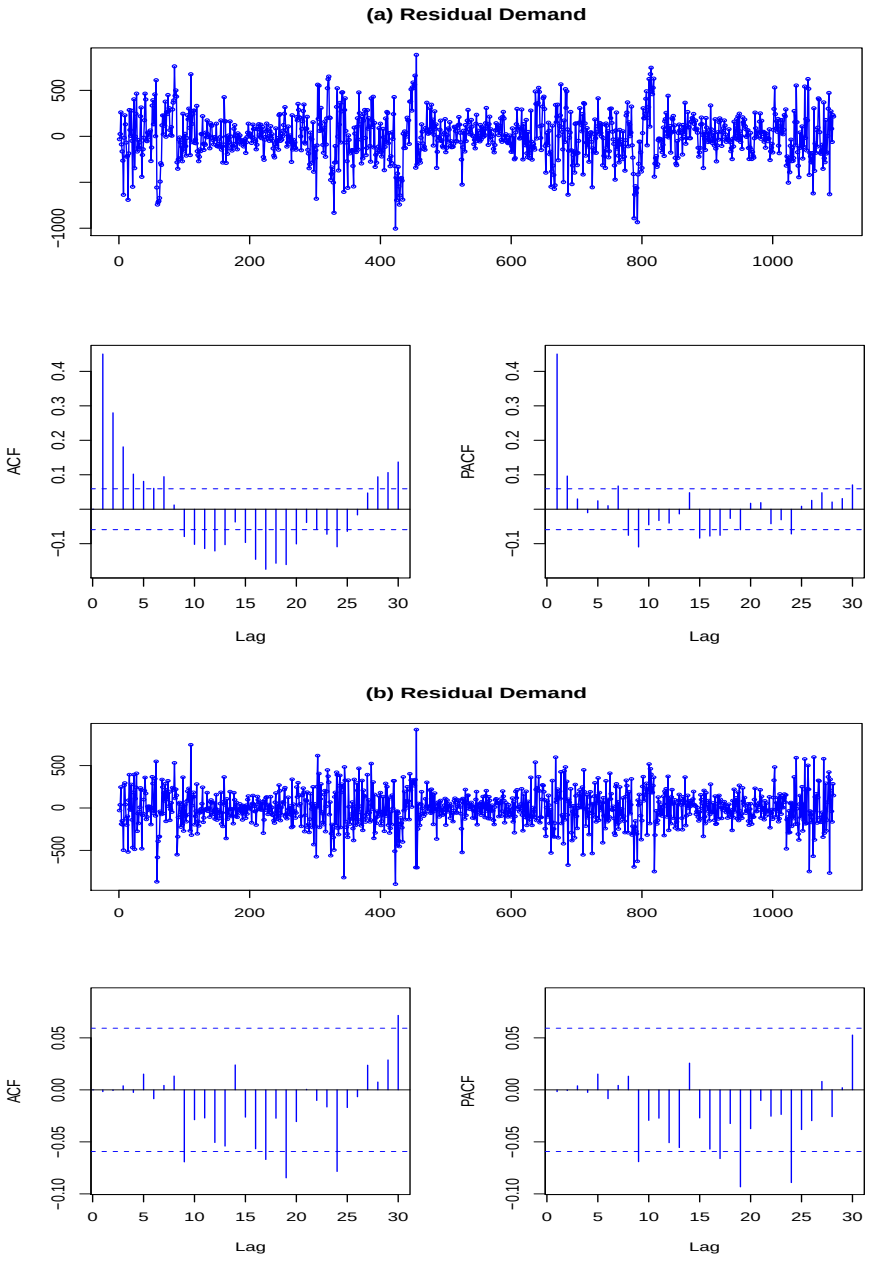

Figure 4.4: **Top**: Time series display for the residuals before reducing autocorrelation for model (4.3.3). **Bottom**: Time series display for the residuals after reducing autocorrelation for model (4.3.3).

4.3.4 Model selection

Table 4.5 presents model selection. The best model is selected using Adjusted R^2, Akaike information criterion (AIC), Bayesian Information Criterion (BIC) and Generalized Cross Validation (GCV). The smallest values of the AIC, BIC and GCV along with the highest Adjusted R^2 represent the best model.

Table 4.5: Model selection for model (4.3.2) and model (4.3.3) for 19:00.

Model	Adjusted R^2	AIC	BIC	GCV
Model (4.3.2)	0.982	16068.07	16216.14	1.4176e+05
Model (4.3.3)	0.992	15235.30	15582.24	66394

4.3.5 Tensor product interactions for 19:00

This section presents tensor product interactions using functions te, ti and $t2$ from the "mgcv" r package (Wood, 2017). The function ti is used for tensor product interactions with main effects and any lower interactions, while te is for a full tensor product smooth. The function $t2$ is an alternative function to te and uses different penalization (Laurinec, 2017 and Wood, 2006). We consider main effects and interactions from predictor variables selected using Lasso via hierarchical interactions from Table 4.4. From model (4.3.3), we replace s by tensor product interaction functions. The tensor product interaction models for hour 19:00 using $te, ti, t2$ and smoothing s are given by:

$$y_{t,19} = \beta_0 + s_{1(19)}(Lag_1) + s_{2(19)}(Lag_2) + s_{3(19)}(\text{DPED}) + s_{4(19)}(\text{minED}) + s_{5(19)}(\text{AveED}) + s_{6(19)}(\text{AminT}) + s_{7(19)}(\text{AveT}) + s_{8(19)}(\text{AmaxT}) + s_{9(19)}(\text{DBH}) + s_{10(19)}(\text{DH}) + s_{11(19)}(\text{DAH}) + s_{12(19)}(\text{Month}) + s_{13(19)}(\text{DayType}) + s_{14(19)}(\text{trend}) + te(\text{Lag}_1,\text{DPED}) + te(\text{Lag}_1,\text{AmaxT}) + te(\text{Lag}_2,\text{DPED}) + te(\text{Lag}_2,\text{AmaxT}) + \xi_{t(19)} \tag{4.3.5}$$

$$y_{t,19} = \beta_0 + s_{1(19)}(Lag_1) + s_{2(19)}(Lag_2) + s_{3(19)}(\text{DPED}) + s_{4(19)}(\text{minED}) + s_{5(19)}(\text{AveED}) + s_{6(19)}(\text{AminT}) + s_{7(19)}(\text{AveT}) + s_{8(19)}(\text{AmaxT}) + s_{9(19)}(\text{DBH}) + s_{10(19)}(\text{DH}) + s_{11(19)}(\text{DAH}) + s_{12(19)}(\text{Month}) + s_{13(19)}(\text{DayType}) + s_{14(19)}(\text{trend}) + ti(\text{Lag}_1,\text{DPED}) + ti(\text{Lag}_1,\text{AmaxT}) + ti(\text{Lag}_2,\text{DPED}) + ti(\text{Lag}_2,\text{AmaxT}) + \xi_{t(19)} \tag{4.3.6}$$

$$y_{t,19} = \beta_0 + S_{1(19)}s(Lag_1) + s_{2(19)}(Lag_2) + s_{3(19)}(\text{DPED}) + s_{4(19)}(\text{minED}) + s_{5(19)}(\text{AveED}) + s_{6(19)}(\text{AminT}) + s_{7(19)}(\text{AveT}) + s_{8(19)}(\text{AmaxT}) + s_{9(19)}(\text{DBH}) + s_{10(19)}(\text{DH}) + s_{11(19)}(\text{DAH}) + s_{12(19)}(\text{Month}) + s_{13(19)}(\text{DayType}) + s_{14(19)}(\text{trend}) + t2(\text{Lag}_1,\text{DPED}) + t2(\text{Lag}_1,\text{AmaxT}) + t2(\text{Lag}_2,\text{DPED}) + t2(\text{Lag}_2,\text{AmaxT}) + \xi_{t(19)} \tag{4.3.7}$$

$$y_{t,19} = \beta_0 + S_{1(19)}s(Lag_1) + s_{2(19)}(Lag_2) + s_{3(19)}(\text{DPED}) + s_{4(19)}(\text{minED}) + s_{5(19)}(\text{AveED}) + s_{6(19)}(\text{AminT}) + s_{7(19)}(\text{AveT}) + s_{8(19)}(\text{AmaxT}) + s_{9(19)}(\text{DBH}) + s_{10(19)}(\text{DH}) + s_{11(19)}(\text{DAH}) + s_{12(19)}(\text{Month}) + s_{13(19)}(\text{DayType}) + s_{14(19)}(\text{trend}) + s_{15(19)}(\text{Lag}_1\text{,DPED}) + s_{16(19)}(\text{Lag}_1\text{,AmaxT}) + s_{17(19)}(\text{Lag}_2\text{,DPED}) + s_{18(19)}(\text{Lag}_2\text{,AmaxT}) + \xi_{t(19)} \tag{4.3.8}$$

From models (4.3.5), (4.3.6), (4.3.7) and (4.3.8), the best model is selected using Adjusted R^2, Akaike information criterion (AIC), Bayesian Information Criterion (BIC) and Generalized Cross Validation (GCV). The smallest values of the AIC, BIC and GCV along with the highest value of Adjusted R^2 represent the best model. Since all models have same values of Adjusted R^2 and nearly similar values of AIC and BIC, we consider the model with lowest value of GCV as best model.

Table 4.6: Model selection for tensor product interactions.

Model	Adjusted R^2	AIC	BIC	GCV
Model (4.3.5)	0.992	15233.73	15566.75	66277
Model (4.3.6)	0.992	15232.50	15555.79	66187
Model (4.3.7)	0.992	15233.34	15549.04	66227
Model (4.3.8)	0.992	15251.19	15657.16	67473

From Table 4.6, the best model is model (4.3.6). Figure (4.5) displays four standard diagnostic plots for the residuals, used for evaluating the fitted best model (4.3.6). From Figure (4.5), the top-left and bottom-left panels

show that the residuals seem to be normally distributed. The histogram also shows that the residuals are left (negatively) skewed. The top-left and bottom-right panels show the plots of the positive residual demand. The best model fit requires the residuals to be normally distributed and no pattern in the predicted values, as look like to be satisfied by model (4.3.6).

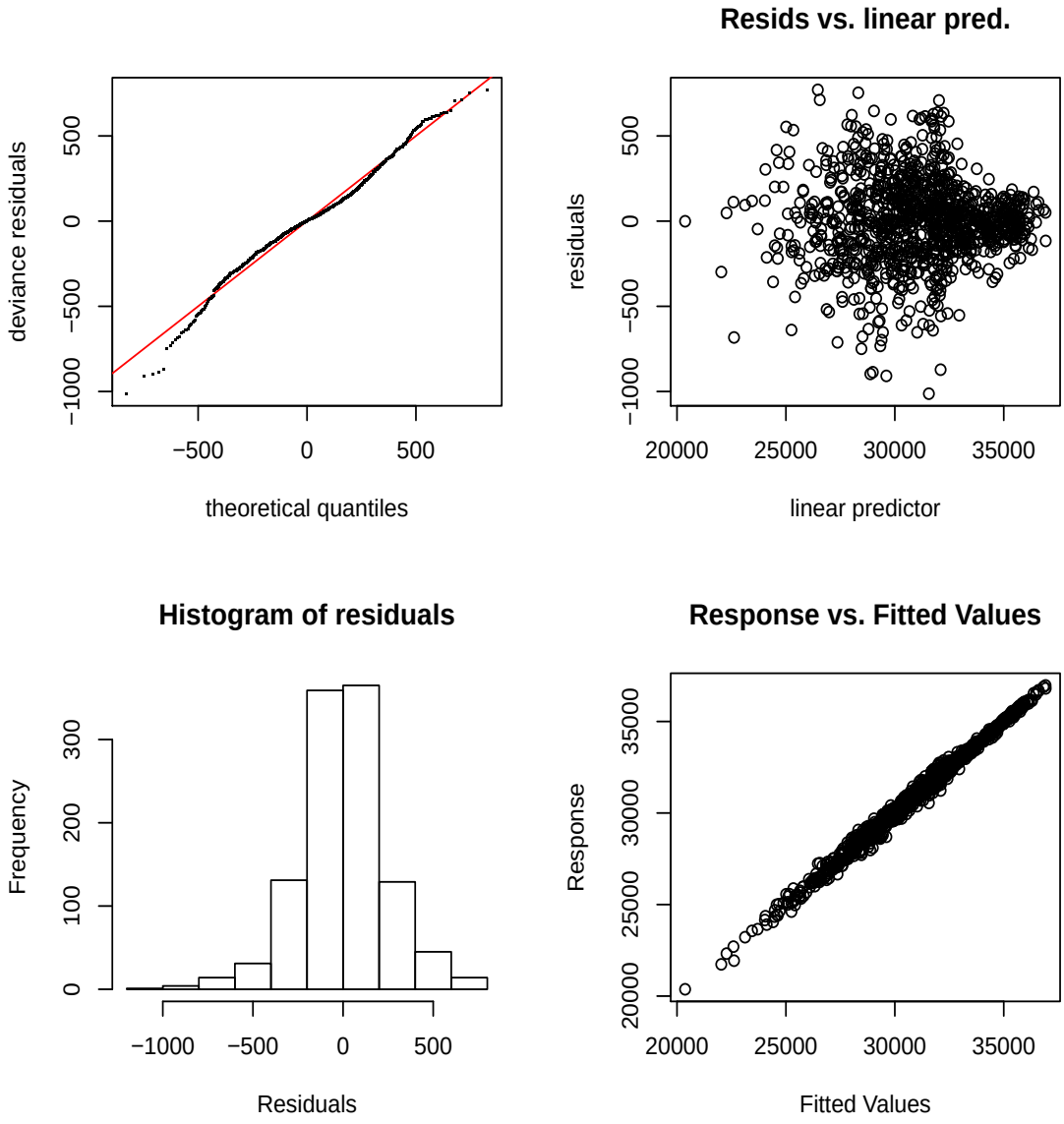

Figure 4.5: Diagnostic plots for the residuals for model (4.3.6). **Top left**: Quantile-Quantile (Q-Q) plot of residuals. **Top right**: linear predictor vs. residuals. **Bottom left**: histogram of residuals. **Bottom right**: the plot of fitted values vs. response.

Since the tensor product function ti is the best function in tensor product interactions for hour 19:00. Tensor product function ti is used to evaluate the relationship between interacting variables (DayType and Month) and demand at hour 19:00. Figure 4.6 display 3D and 2D contour representation of ti(DayType,Month). The idea of Figure 4.6 is to bring understanding on how interacting variables in tensor product affect load demand.

From Figure (4.6)a. the highest peak is when the DayType variable has values from 1 to 4 (i.e. from Monday to Thursday) and the Month variable has values from 5 to 7 (i.e. from late May to late July). From Figure (4.6)b. the highest value of electricity load is on Monday, it is very similar till Thursday, then the load is decreasing on weekends. Since big industries open from Monday to Friday, the big machines which require huge electricity to operate are used. Thus the reason for high demand from Monday to Thursday.

On Month variable the highest value of electricity load is on late May, it is very similar till late July, then the load is decreasing from late August. The electricity load is increasing from January till peak demand on late May. Many households and industries used more electricity in June and July because, on winter the industries cannot generate their own electricity due to low temperature. Also households use many electricity appliances (i.e. heater, stove, etc.) at the same time due to low temperature.

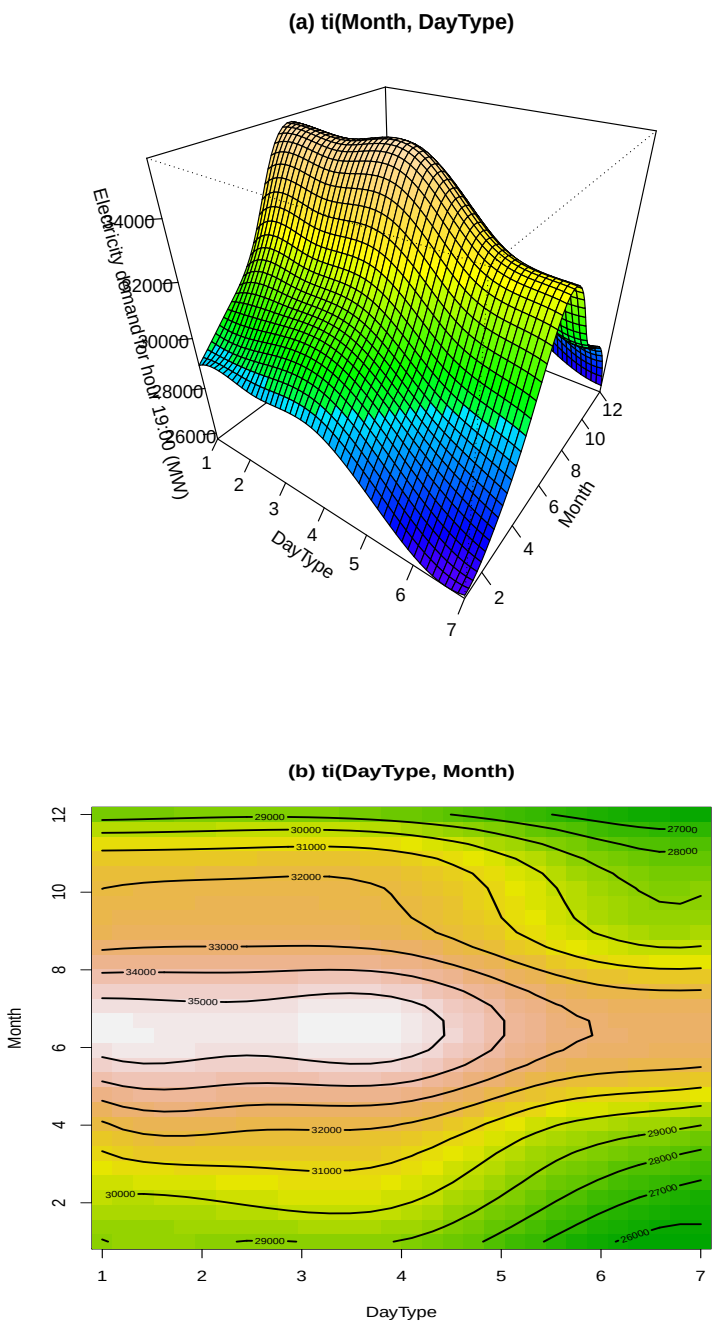

Figure 4.6: 3D and 2D contour for ti(DayType,Month).

4.3.6 Evaluation of forecasts from best models with and without tensor product interactions

Mean absolute percentage error (MAPE), mean absolute error (MAE) and root mean square error (RMSE) are used to evaluate the forecasting accuracy of best prediction models. We consider best model (4.3.3) without tensor product interactions and best model (4.3.6) with tensor product interactions. The model with the smallest MAPE, MAE and RMSE is the best model, which is model (4.3.6). Table 4.7 displays accuracy measures for evaluation of forecasts during the testing period.

Table 4.7: Evaluation of forecasts from best models with and without tensor product interactions for 19:00.

Performance criteria	Model (4.3.3)	Model (4.3.6)
MAPE	0.68	0.67
MAE	204.58	200.07
RMSE	263.38	261.05

4.3.7 Forecasting results for 19:00

The training period for the load forecasting model is from January 2009 to December 2011. To evaluate the proposed best model (4.3.6) performance we used the testing period which is from January 2012 to December 2013. Comparatively, it could be seen in the top panel of Figure 4.7 that the forecasts follow the actual demand remarkably well. The bottom panel of Figure 4.7 displays the densities of the forecasts for the best GAM (4.3.6) for the

period January 2012 to December 2013.

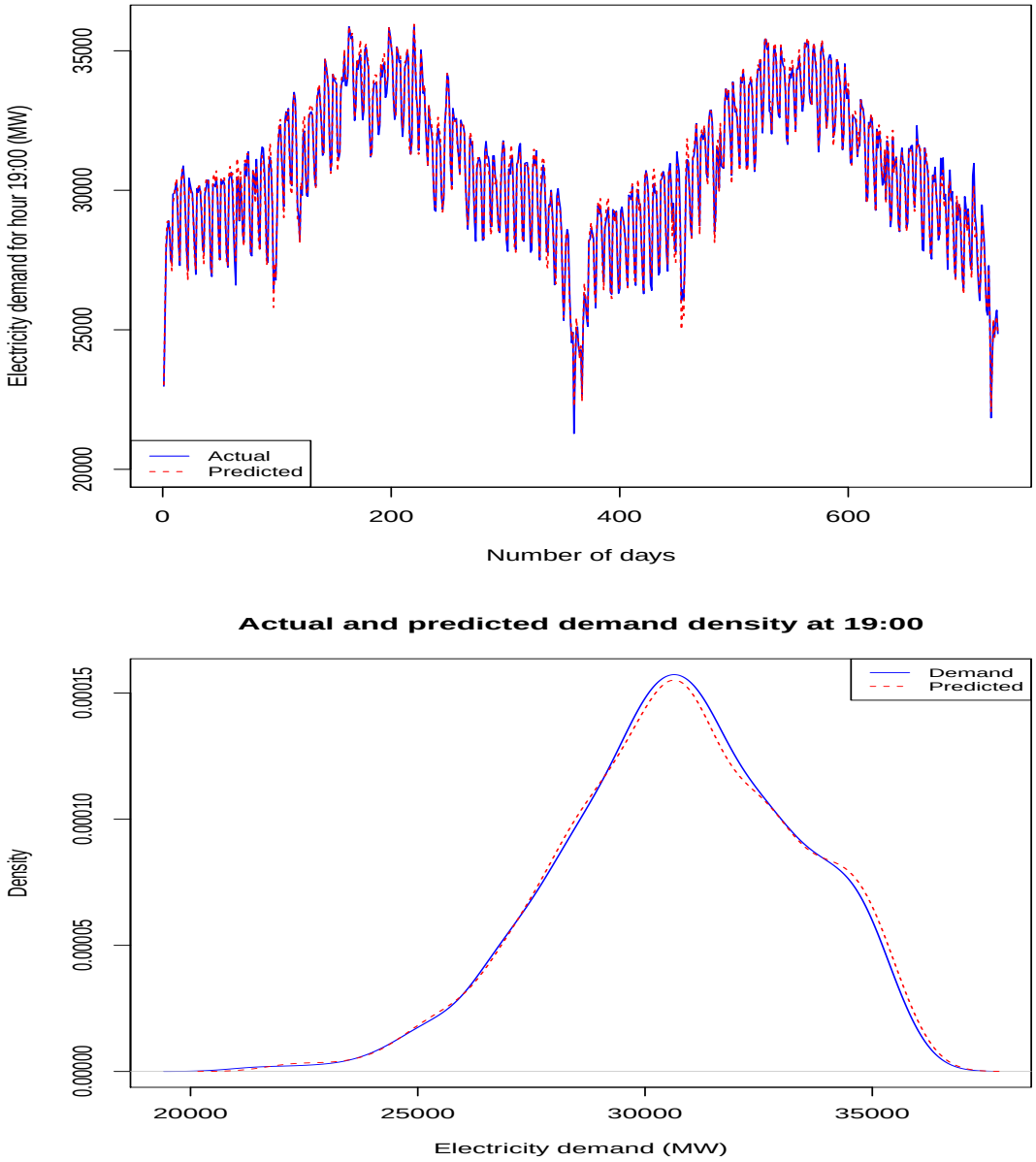

Figure 4.7: **Top panel**: Actual and predicted electricity demand in megawatts at hour 19:00 for January 2012 to December 2013. **Bottom panel**: Kernel density of the actual and predicted electricity demand at hour 19:00 for January 2012 to December 2013.

4.4 Generalized additive model for 20:00

4.4.1 Introduction

Similarly with hour 19:00, the different features of the models which include variable description, variable selection and model building are discussed in this section.

4.4.2 Variable and model selection

A simple generalized additive model (GAM) which include all the predictor variables is considered before any variable selection using Lasso and Lasso via hierarchical interactions. The first GAM with main effects for 20:00 is given by:

$$\begin{aligned} y_{t,20} = \beta_0 + s_{1(20)}(Lag_1) + s_{2(20)}(Lag_2) + s_{3(20)}(\text{DPED}) + s_{4(20)}(\text{minED}) + \\ s_{5(20)}(\text{AveED}) + s_{6(20)}(\text{AminT}) + s_{7(20)}(\text{AveT}) + s_{8(20)}(\text{AmaxT}) + \\ s_{9(20)}(\text{DBH}) + s_{10(20)}(\text{DH}) + s_{11(20)}(\text{DAH}) + s_{12(20)}(\text{Month}) + \\ s_{13(20)}(\text{DayType}) + s_{14(20)}(\text{trend}) + \xi_{t(20)} \end{aligned} \tag{4.4.1}$$

where $y_{t,20}$, represent the electricity demand on day t at hour 20:00. β_0 and $s_{1(20)}, ..., s_{14(20)}$ are parameters to be estimated.

Variable selection using Lasso for 20:00

The response variable $y_{t,20}$ and all predictor variables in model (4.4.1) including Lag_1 and Lag_2, are used for Lasso variables selections. Table 4.8 shows variables selection using Lasso for hour 20:00. From Table 4.8 the variables

Table 4.8: Variable selected using Lasso for 20:00 hour demand.

(Intercept)	2019.40389226
Lag_1	0.12609617
Lag_2	0.01260085
DPED	0.71727444
minED	0.08686109
AveED	0.15929433
AminT	0
AveT	0
AmaxT	0
DBH	0
DH	-231.63159071
DAH	0
Month	9.24122047
DayType	0
trend	0.15548339

with zero coefficients are not selected. Variables selected are Lag_1, Lag_2, DPED, minED, AveED, DH, Month and trend. GAM after Lasso variable selection is given by:

$$y_{t,20} = \beta_0 + s_{1(20)}(Lag_1) + s_{2(20)}(Lag_2) + s_{3(20)}(\text{DPED}) + s_{4(20)}(\text{minED}) + s_{5(20)}(\text{AveED}) + s_{6(20)}(\text{DH}) + s_{7(20)}(\text{Month}) + s_{8(20)}(\text{trend}) + \xi_{t(20)} \tag{4.4.2}$$

Variable selection using Lasso via hierarchical interactions for 20:00.

Table 4.4.3 shows variable selection using Lasso via hierarchical interactions for 20:00, as we did for hour 19:00.

Table 4.9: Variable selected using Lasso via hierarchical interactions for 20:00 hour demand

	Main effect	Lag_1	Lag_2	Month	Tight?
Lag_1	210.5241	-62.9782	0	0	
DPED	1592.4653	-3.7803	0	32.587	
AminT	-8.3539	0	0.6965	0	*
AmaxT	2.2742	0	2.2742	0	*
DH	-15.1648	0	0	-4.3745	*
trend	58.6044	0	0	-7.1409	

From Table 4.4.3, interacting variables are Lag_1 interact with DPED, Lag_2 interact with AminT, Lag_2 interact with AmaxT, DPED interact with Month, DH interact with Month and Month interact with trend. The GAM after variable selection using Lasso via hierarchical interactions for 20:00 is given by:

$$\begin{aligned} y_{t,20} = \beta_0 + s_{1(20)}(Lag_1) + s_{2(20)}(Lag_2) + s_{3(20)}(\text{DPED}) + s_{4(20)}(\text{minED}) + \\ s_{5(20)}(\text{AveED}) + s_{6(20)}(\text{AminT}) + s_{7(20)}(\text{AveT}) + s_{8(20)}(\text{AmaxT}) + \\ s_{9(20)}(\text{DBH}) + s_{10(20)}(\text{DH}) + s_{11(20)}(\text{DAH}) + s_{12(20)}(\text{Month}) + s_{13(20)}(\text{DayType}) + \\ s_{14(20)}(\text{trend}) + s_{15(20)}(\text{Lag}_1,\text{DPED}) + s_{16(20)}(\text{Lag}_2,\text{AminT}) + s_{17(20)}(\text{Lag}_2,\text{AmaxT}) + \\ s_{18(20)}(\text{DPED,Month}) + s_{19(20)}(\text{DH,Month}) + s_{19(20)}(\text{Month,trend}) + \xi_{t(20)} \end{aligned} \tag{4.4.3}$$

4.4.3 Model diagnostics

GAM (4.4.2) is used for checking autocorrelation. After testing for autocorrelation, the p-value of residuals is 0.00 (p=0.00). Since the p-value is less than 0.05, we conclude that there is autocorrelations in the residuals. Furthermore we reduce the autocorrelation, to make p-value greater than 0.05. In order to reduce the autocorrelation, SARIMA$(5,0,2)(2,0,3)_7$ (given by Eq(4.3.4)) is used and the autocorrelation is reduced until getting p-value equal to 0.90 (p=0.90). Since p-value is greater than 0.05, we conclude that there is no autocorrelation in the residuals at 5% level of significance.

Secondly the GAM (4.4.3), with the main effects and interactions is considered to check the residuals autocorrelation. After testing for autocorrelation, the p-value of residuals is 0.00 (p=0.00). Since the p-value is less than 0.05, we conclude that there is autocorrelations in the residuals. Therefore we reduce the autocorrelation, to make p-value greater than 0.05.

In order to reduce the autocorrelation, SARIMA$(5,0,2)(2,0,3)_7$ (given by Eq(4.3.4)) is used and the autocorrelation is reduced until getting p-value equal to 0.99 (p=0.99). Since p-value is greater than 0.05, we conclude that there is no autocorrelation in the residuals at 5% level of significance. Time series display of residuals for GAMs (4.4.2) and (4.4.3) before and after correcting autocorrelation are given in Figure 4.8 and 4.9 respectively.

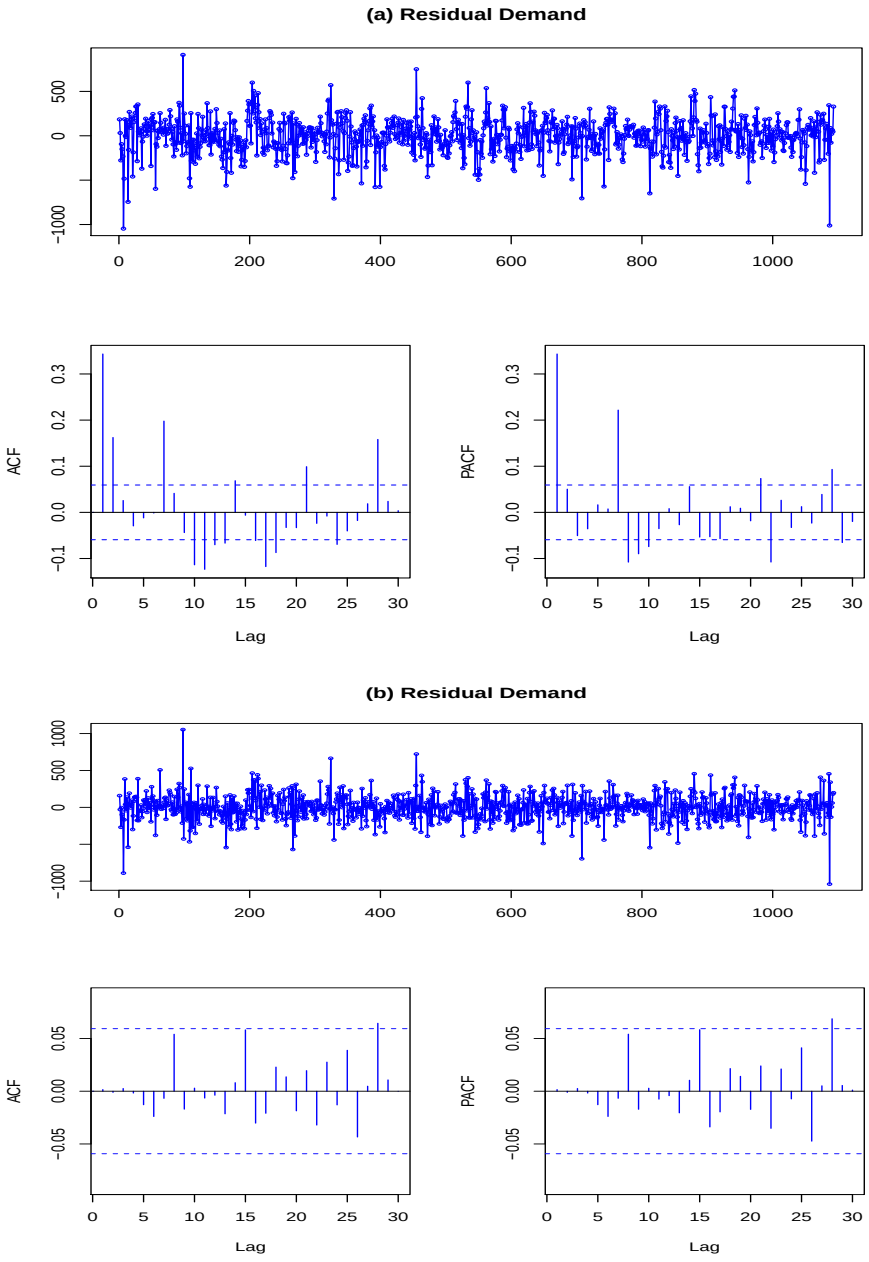

Figure 4.8: **Top**: Time series display for the residuals before reducing autocorrelation for model (4.4.2). **Bottom**: Time series display for the residuals after reducing autocorrelation for model (4.4.2).

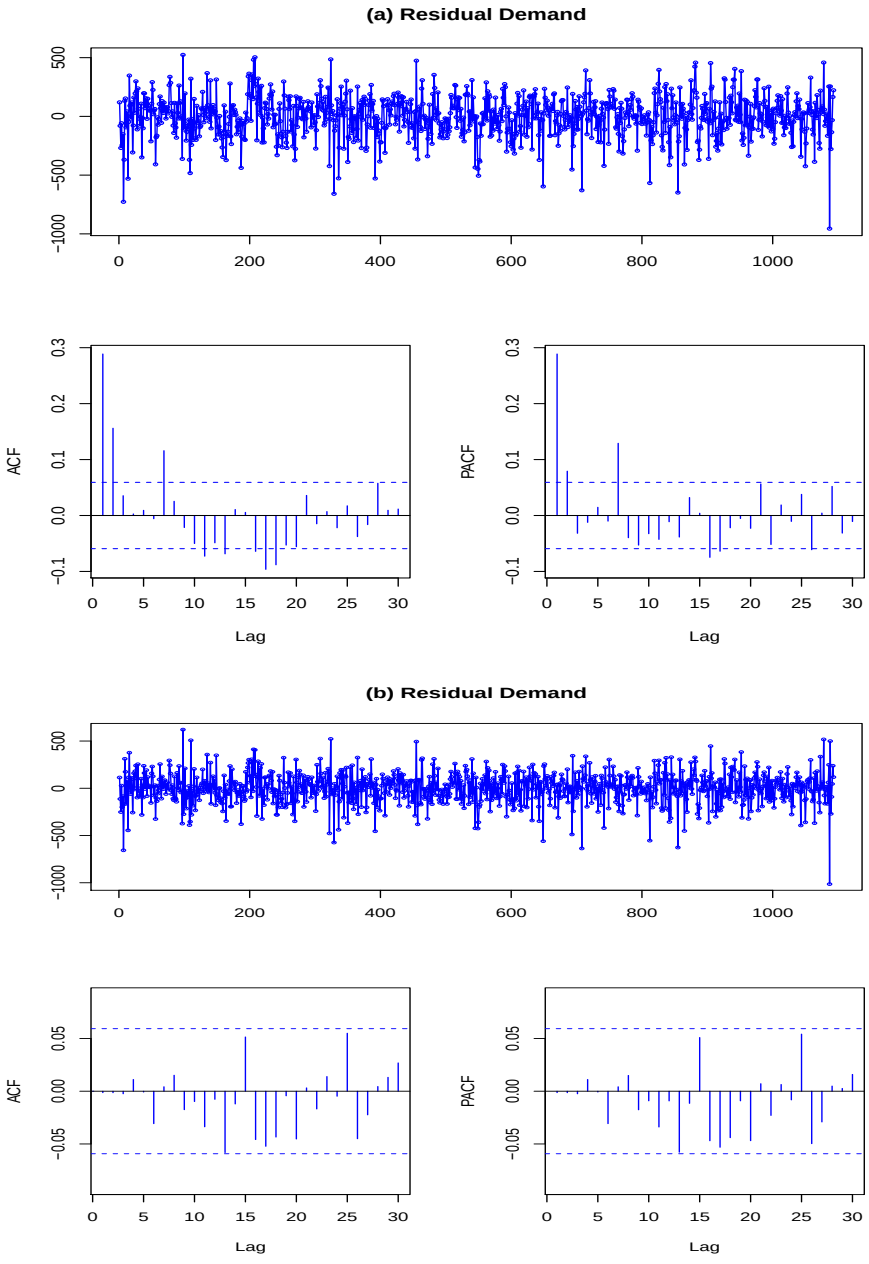

Figure 4.9: **Top**: Time series display for the residuals before reducing autocorrelation for model (4.4.3). **Bottom**: Time series display for the residuals after reducing autocorrelation for model (4.4.3).

4.4.4 Model selection

Table 4.10 presents model selection. The best model is selected using Adjusted R^2, Akaike information criterion (AIC), Bayesian Information Criterion (BIC) and Generalized Cross Validation (GCV). The minimum values of the AIC, BIC and GCV along with the highest Adjusted R^2 value represent the best model.

Table 4.10: Model selection for model (4.4.2) and model (4.4.3) for 20:00.

Model	Adjusted R^2	AIC	BIC	GCV
Model (4.4.2)	0.991	14777.53	15061.50	43615
Model (4.4.3)	0.993	14547.41	15055.35	35559

4.4.5 Tensor product interactions for 20:00

The tensor product interaction models for hour 20:00 using $te, ti, t2$ and smoothing s are given by:

$$\begin{aligned} y_{t,20} = \beta_0 &+ s_{1(20)}(Lag_1) + s_{2(20)}(Lag_2) + s_{3(20)}(\text{DPED}) + s_{4(20)}(\text{minED}) + \\ &s_{5(20)}(\text{AveED}) + s_{6(20)}(\text{AminT}) + s_{7(20)}(\text{AveT}) + s_{8(20)}(\text{AmaxT}) + \\ &s_{9(20)}(\text{DBH}) + s_{10(20)}(\text{DH}) + s_{11(20)}(\text{DAH}) + s_{12(20)}(\text{Month}) + s_{13(20)}(\text{DayType}) + \\ &s_{14(20)}(\text{trend}) + te(\text{Lag}_1,\text{DPED}) + te(\text{Lag}_2,\text{AminT}) + \\ &te(\text{Lag}_2,\text{AmaxT}) + te(\text{DPED},\text{Month}) + te(\text{Month},\text{trend}) + \xi_{t(20)} \end{aligned} \tag{4.4.4}$$

$$y_{t,20} = \beta_0 + s_{1(20)}(Lag_1) + s_{2(20)}(Lag_2) + s_{3(20)}(\text{DPED}) + s_{4(20)}(\text{minED}) + s_{5(20)}(\text{AveED}) + s_{6(20)}(\text{AminT}) + s_{7(20)}(\text{AveT}) + s_{8(20)}(\text{AmaxT}) + s_{9(20)}(\text{DBH}) + s_{10(20)}(\text{DH}) + s_{11(20)}(\text{DAH}) + s_{12(20)}(\text{Month}) + s_{13(20)}(\text{DayType}) + s_{14(20)}(\text{trend}) + ti(\text{Lag}_1,\text{DPED}) + ti(\text{Lag}_2,\text{AminT}) + ti(\text{Lag}_2,\text{AmaxT}) + ti(\text{DPED},\text{Month}) + ti(\text{Month},\text{trend}) + \xi_{t(20)} \quad (4.4.5)$$

$$y_{t,20} = \beta_0 + s_{1(20)}(Lag_1) + s_{2(20)}(Lag_2) + s_{3(20)}(\text{DPED}) + s_{4(20)}(\text{minED}) + s_{5(20)}(\text{AveED}) + s_{6(20)}(\text{AminT}) + s_{7(20)}(\text{AveT}) + s_{8(20)}(\text{AmaxT}) + s_{9(20)}(\text{DBH}) + s_{10(20)}(\text{DH}) + s_{11(20)}(\text{DAH}) + s_{12(20)}(\text{Month}) + s_{13(20)}(\text{DayType}) + s_{14(20)}(\text{trend}) + t2(\text{Lag}_1,\text{DPED}) + t2(\text{Lag}_2,\text{AminT}) + t2(\text{Lag}_2,\text{AmaxT}) + t2(\text{DPED},\text{Month}) + t2(\text{Month},\text{trend}) + \xi_{t(20)} \quad (4.4.6)$$

$$y_{t,20} = \beta_0 + s_{1(20)}(Lag_1) + s_{2(20)}(Lag_2) + s_{3(20)}(\text{DPED}) + s_{4(20)}(\text{minED}) + s_{5(20)}(\text{AveED}) + s_{6(20)}(\text{AminT}) + s_{7(20)}(\text{AveT}) + s_{8(20)}(\text{AmaxT}) + s_{9(20)}(\text{DBH}) + s_{10(20)}(\text{DH}) + s_{11(20)}(\text{DAH}) + s_{12(20)}(\text{Month}) + s_{13(20)}(\text{DayType}) + s_{14(20)}(\text{trend}) + s_{15(20)}(\text{Lag}_1,\text{DPED}) + s_{16(20)}(\text{Lag}_2,\text{AminT}) + s_{17(20)}(\text{Lag}_2,\text{AmaxT}) + s_{18(20)}(\text{DPED},\text{Month}) + s_{19(20)}(\text{Month},\text{trend}) + \xi_{t(20)} \quad (4.4.7)$$

From models (4.4.4), (4.4.5), (4.4.6) and (4.4.7), the best model is selected using Adjusted R^2, Akaike information criterion (AIC), Bayesian Information Criterion (BIC) and Generalized Cross Validation (GCV). The smallest values of the AIC, BIC and GCV along with the highest value of Adjusted R^2 represent the best model. Since all models has same values of Adjusted R^2 and nearly similar values of AIC and BIC, we consider the model with lowest value of GCV as best model.

Table 4.11: Model selection for tensor product interactions.

Model	Adjusted R^2	AIC	BIC	GCV
Model (4.4.4)	0.993	14492.79	15105.39	33973
Model (4.4.5)	0.993	14543.32	14991.78	35353
Model (4.4.6)	0.993	14554.95	14935.33	35659
Model (4.4.7)	0.993	14551.97	15097.05	35759

From Table 4.11, the best model is model (4.4.4). Figure 4.10 displays four standard diagnostic plots for the residuals, used for evaluating the fitted best model (4.4.4). From Figure 4.10, the top-left and bottom-left panels show that the residuals are seems to be normally distributed. Histogram also shows that the residuals is left (negatively) skewed. The top-left and bottom-right panels show the plots of the positive residual demand. The best model fit requires the residuals to be normally distributed and no pattern in the predicted values, as look like to be satisfied by model (4.4.4).

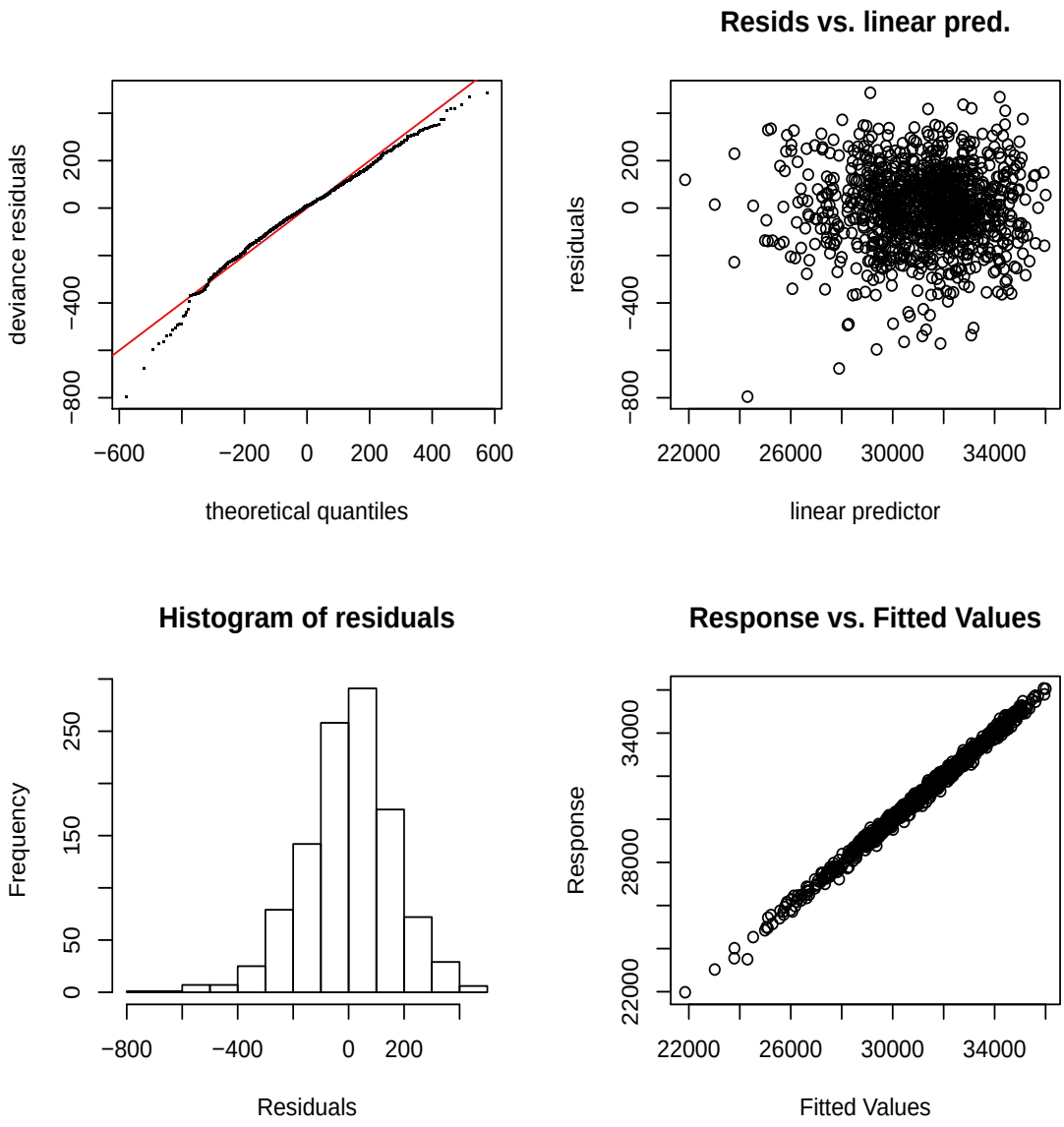

Figure 4.10: Diagnostic plots for the residuals for model (4.4.4). **Top left**: Quantile-Quantile (Q-Q) plot of residuals. **Top right**: linear predictor vs. residuals. **Bottom left**: histogram of residuals. **Bottom right**: the plot of fitted values vs. response.

Since the tensor product function *te* is the best function in tensor product interactions for hour 20:00. Tensor product function *te* is used to evaluate the relationship between interacting variables (DayType and Month) and demand at hour 20:00. Figure 4.11 display 3D and 2D contour representation of *te*(DayType,Month).

From Figure (4.6)a. the highest peak is when the DayType variable has values from 1 to 4 (i.e. from Monday to Thursday) and the Month variable has values from 5 to 7 (i.e. from late May to late July). From Figure (4.11)b. the highest value of electricity load is on Monday, it is very similar till Thursday, then the load is decreasing on weekends. On Month variable the highest value of electricity load is on late May, it is very similar till late July, then the load is decreasing from late August. The electricity load is increasing from January till peak demand on late May.

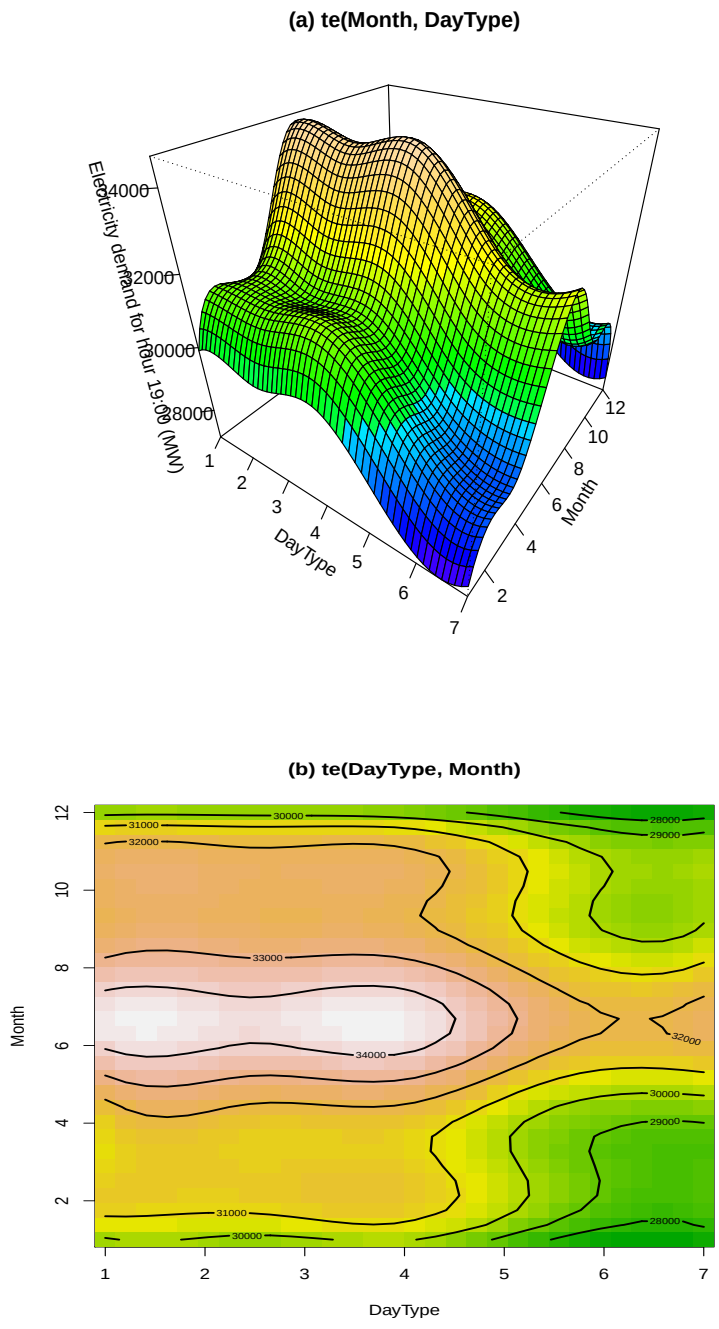

Figure 4.11: 3D and 2D contour for te(DayType,Month).

4.4.6 Evaluation of forecasts from best models with and without tensor product interactions

MAPE, MAE and RMSE are used to evaluate the forecasting accuracy of best prediction models. We consider best model (4.4.3) without tensor product interactions and best model (4.4.4) with tensor product interactions. The model with the smaller MAPE, MAE and RMSE is the best model, which is model (4.4.3). Table 4.12 displays accuracy measures for evaluation of forecasts during testing period.

Table 4.12: Evaluation of forecasts from best models with and without tensor product interactions for 20:00.

Performance criteria	Model (4.4.3)	Model (4.4.4)
MAPE	1.08	5.49
MAE	335.61	1669.55
RMSE	410.22	1911.27

4.4.7 Forecasting results for 20:00

To evaluate the proposed best model (4.4.3) performance we used the testing period which is from January 2012 to December 2013. Comparatively, it could be seen in the top panel of Figure 4.12 that the forecasts follow the actual demand remarkably well. The bottom panel of Figure 4.12 displays the densities plot of the forecasts for the best GAM (4.4.3) for the period January 2012 to December 2013.

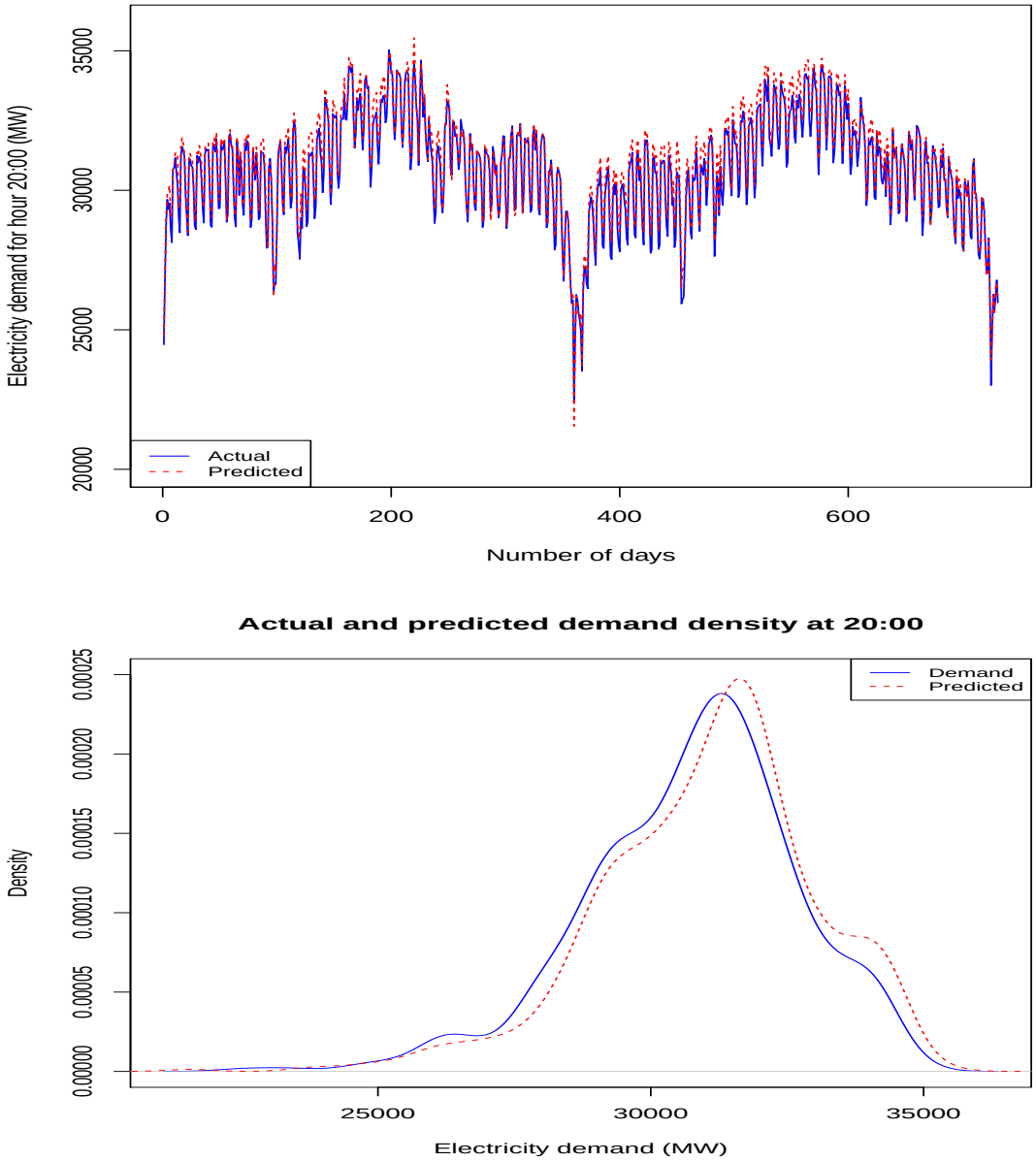

Figure 4.12: **Top panel**: Actual and predicted electricity demand in megawatts at hour 20:00 for January 2012 to December 2013. **Bottom panel**: Kernel density of the actual and predicted electricity demand at hour 20:00 for January 2012 to December 2013.

4.5 Generalized additive model for DPED

4.5.1 Introduction

This section present different features of the models which include variable description, variable selection and model building for daily peak electricity demand (DPED).

4.5.2 Variable and model selection

A simple generalized additive model (GAM) which include all the predictor variables is considered before any variable selection using Lasso and Lasso via hierarchical interactions. The first GAM with main effects for DPED is given by:

$$\begin{aligned} y_{t,DP} = \beta_0 + s_{1(DP)}(Lag_1) + s_{2(DP)}(Lag_2) + s_{3(DP)}(\text{DPED}) + s_{4(DP)}(\text{minED}) + \\ s_{5(DP)}(\text{AveED}) + s_{6(DP)}(\text{AminT}) + s_{7(DP)}(\text{AveT}) + s_{8(DP)}(\text{AmaxT}) + \\ s_{9(DP)}(\text{DBH}) + s_{10(DP)}(\text{DH}) + s_{11(DP)}(\text{DAH}) + s_{12(DP)}(\text{Month}) + \\ s_{13(DP)}(\text{DayType}) + s_{14(DP)}(\text{trend}) + \xi_{t(DP)} \end{aligned} \tag{4.5.1}$$

where $y_{t,DP}$, represent the daily peak electricity demand on day t. β_0 and $s_{1(20)}, ..., s_{14(20)}$ are parameters to be estimated.

Variable selection using Lasso for DPED

The response variable $y_{t,DP}$ and all predictor variables in model (4.5.1) including Lag_1 and Lag_2, are used for Lasso variables selections. Table 4.13 shows variables selection using Lasso for hour DPED.

Table 4.13: Variable selected using Lasso for DPED

(Intercept)	2873.34400161
Lag_1	0.08379628
Lag_2	0
minED	-0.74071099
AveED	1.66140407
AminT	-24.06335195
AveT	-37.30426333
AmaxT	0
DBH	-120.95692521
DH	49.97866770
DAH	0
Month	-5.36283000
DayType	147.79365479
trend	0.06453108

From Table 4.13 the variables with zero coefficients are not selected. Variables selected are Lag_1, minED, AveED, AminT, AveT, DBH, DH, Month, DayType and trend. GAM after Lasso variable selection is given by:

$$y_{t,DP} = \beta_0 + s_{1(DP)}(Lag_1) + s_{2(DP)}(\text{minED}) + s_{3(DP)}(\text{AveED}) + s_{4(DP)}(\text{AminT}) + s_{5(DP)}(\text{AveT}) + s_{6(DP)}(\text{DBH}) + s_{7(DP)}(\text{DH}) + s_{8(DP)}(\text{Month}) + s_{9(DP)}(\text{DayType}) + s_{10(DP)}(\text{trend}) + \xi_{t(DP)} \quad (4.5.2)$$

Variable selection using Lasso via hierarchical interactions for DPED

Table 4.14 shows variable selection using Lasso via hierarchical interactions for DPED, as we did for hour 19:00 and 20:00.

Table 4.14: Variable selected using LASSO via hierarchical interactions for DPED hour demand

	Main effect	AveED	DH	Tight?
Lag_1	285.7314	4.7989	0	
Lag_2	41.2935	39.9037	1e-04	
DayType	48.1252	0	-13.5069	

From Table 4.14, interacting variables are Lag_1 interact with AveED, Lag_2 interacting with AveED, Lag_2 interacting with DH, and DayType interacting with DH. The GAM after variable selection using Lasso via hierarchical interactions for DPED is given by:

$$y_{t,DP} = \beta_0 + s_{1(DP)}(Lag_1) + s_{2(DP)}(Lag_2) + s_{3(DP)}(\text{DPED}) + s_{4(DP)}(\text{minED}) + s_{5(DP)}(\text{AveED}) + s_{6(DP)}(\text{AminT}) + s_{7(DP)}(\text{AveT}) + s_{8(DP)}(\text{AmaxT}) + s_{9(DP)}(\text{DBH}) + s_{10(DP)}(\text{DH}) + s_{11(DP)}(\text{DAH}) + s_{12(DP)}(\text{Month}) + s_{13(DP)}(\text{DayType}) + s_{14(DP)}(\text{trend}) + s_{15(DP)}(\text{Lag}_1,\text{AveED}) + s_{16(DP)}(\text{Lag}_2,\text{AveED}) + s_{17(DP)}(\text{Lag}_2,\text{DH}) + s_{18(DP)}(\text{DayType,DH}) + \xi_{t(DP)} \quad (4.5.3)$$

4.5.3 Model diagnostics

GAM (4.5.2) is used for checking autocorrelation. After testing for autocorrelation, the p-value of residuals is 0.00 (p=0.00). Since the p-value is less than 0.05, we conclude that there is autocorrelations in the residuals. Furthermore we reduce the autocorrelation, to make p-value greater than 0.05. In order to reduce the autocorrelation, SARIMA$(5,0,2)(2,0,3)_7$ (given by Eq(4.3.4)) is used and the autocorrelation is reduced until getting p-value equal to 0.27 (p=0.27). Since p-value is greater than 0.05, we conclude that there is no autocorrelation in the residuals at 5% level of significance.

Secondly the GAM (4.5.3), with the main effects and interactions is considered to check the residuals autocorrelation. After testing for autocorrelation, the p-value of residuals is 0.00 (p=0.00). Since the p-value is less than 0.05, we conclude that there is autocorrelations in the residuals. Therefore we reduce the autocorrelation, to make p-value greater than 0.05.

In order to reduce the autocorrelation, SARIMA$(5,0,2)(2,0,3)_7$ (given by Eq(4.3.4)) is used and the autocorrelation is reduced until getting p-value equal to 1.00 (p=1.00). Since p-value is greater than 0.05, we conclude that there is no autocorrelation in the residuals at 5% level of significance. Time series display of residuals for GAMs (4.5.2) and (4.14) before and after correcting autocorrelation are given in Figure 4.13 and 4.14 respectively.

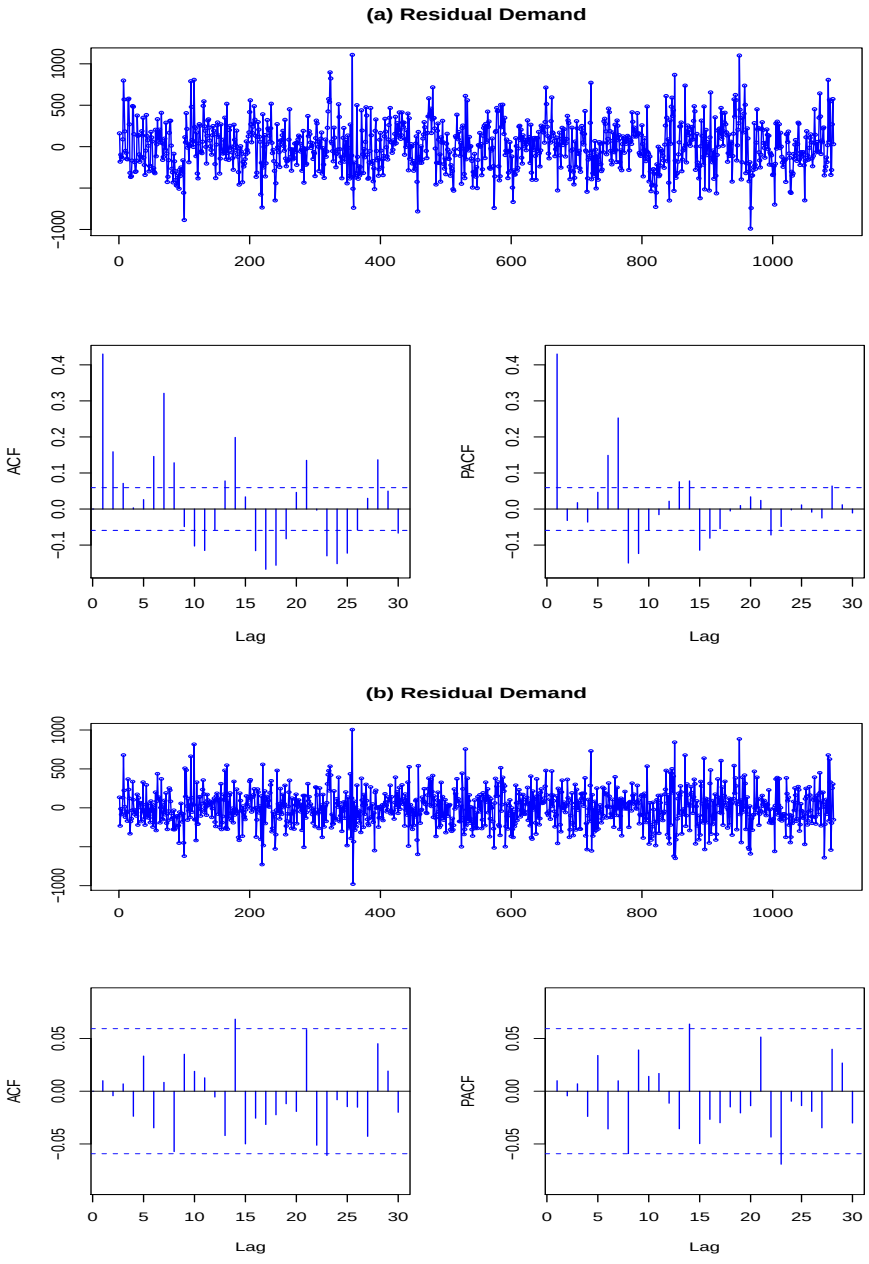

Figure 4.13: **Top**: Time series display for the residuals before reducing autocorrelation for model (4.5.2). **Bottom**: Time series display for the residuals after reducing autocorrelation for model (4.5.2).

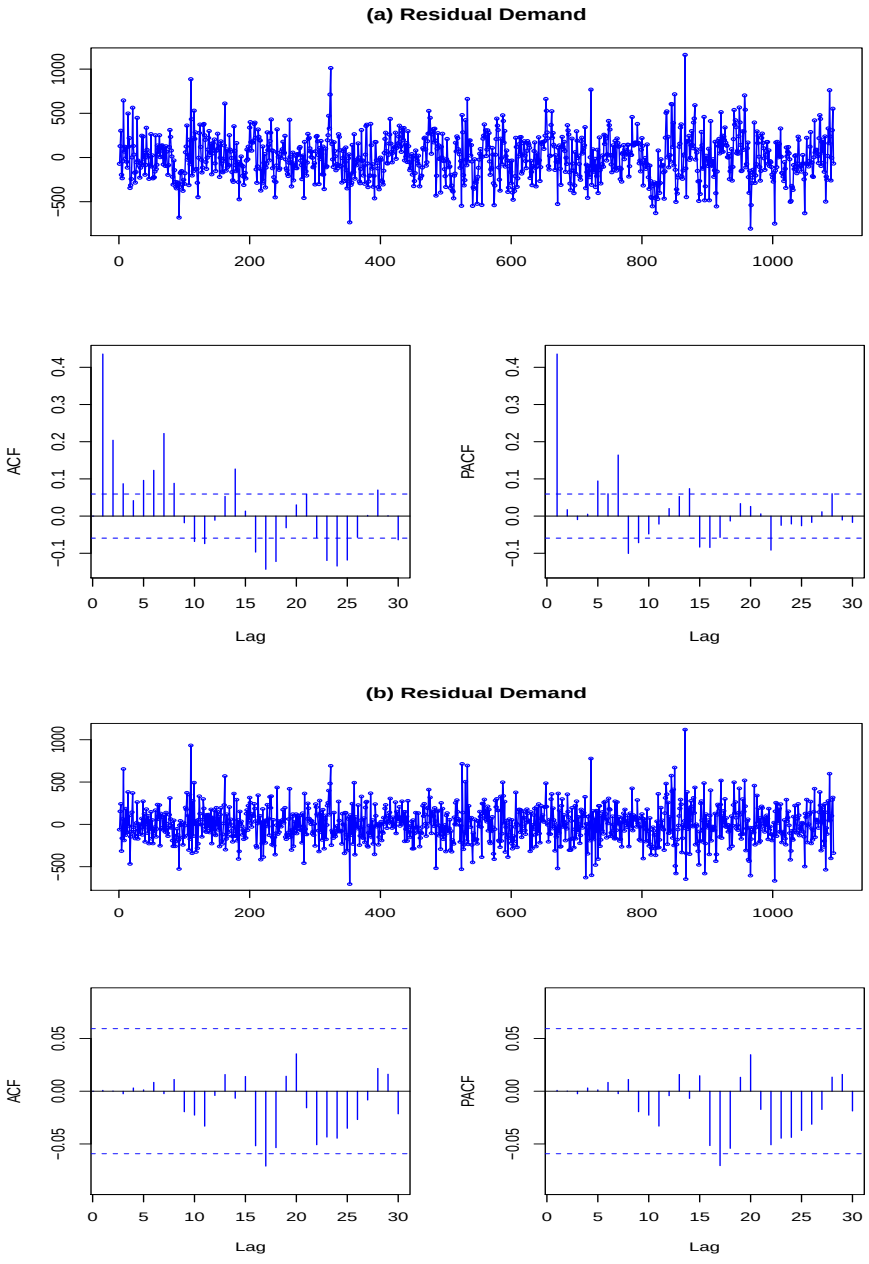

Figure 4.14: **Top**: Time series display for the residuals before reducing autocorrelation for model (4.5.3). **Bottom**: Time series display for the residuals after reducing autocorrelation for model (4.5.3).

4.5.4 Model selection

Table 4.17 presents model selection. The best model is selected using Adjusted R^2, Akaike information criterion (AIC), Bayesian Information Criterion (BIC) and Generalized Cross Validation (GCV). The lowest values of the AIC, BIC and GCV along with the highest Adjusted R^2 value represent the best model.

Table 4.15: Model selection for model (4.5.2) and model (4.5.3) for DPED

Model	Adjusted R^2	AIC	BIC	GCV
Model (4.5.2)	0.985	15475.63	15714.57	82540
Model (4.5.3)	0.988	15237.26	15653.51	66638

4.5.5 Tensor product interactions for DPED

The tensor product interaction models for hour DPED using $te, ti, t2$ and smoothing s are given by:

$$\begin{aligned} y_{t,DP} = \beta_0 &+ s_{1(DP)}(Lag_1) + s_{2(DP)}(Lag_2) + s_{3(DP)}(\text{DPED}) + s_{4(DP)}(\text{minED}) + \\ &s_{5(DP)}(\text{AveED}) + s_{6(DP)}(\text{AminT}) + s_{7(DP)}(\text{AveT}) + s_{8(DP)}(\text{AmaxT}) + \\ &s_{9(DP)}(\text{DBH}) + s_{10(DP)}(\text{DH}) + s_{11(DP)}(\text{DAH}) + s_{12(DP)}(\text{Month}) + s_{13(DP)}(\text{DayType}) + \\ &s_{14(DP)}(\text{trend}) + te(\text{Lag}_1,\text{AveED}) + te(\text{Lag}_2,\text{AveED}) + \xi_{t(DP)} \end{aligned} \tag{4.5.4}$$

$$y_{t,DP} = \beta_0 + s_{1(DP)}(Lag_1) + s_{2(DP)}(Lag_2) + s_{3(DP)}(\text{DPED}) + s_{4(DP)}(\text{minED}) + s_{5(DP)}(\text{AveED}) + s_{6(DP)}(\text{AminT}) + s_{7(DP)}(\text{AveT}) + s_{8(DP)}(\text{AmaxT}) + s_{9(DP)}(\text{DBH}) + s_{10(DP)}(\text{DH}) + s_{11(DP)}(\text{DAH}) + s_{12(DP)}(\text{Month}) + s_{13(DP)}(\text{DayType}) + s_{14(DP)}(\text{trend}) + ti(\text{Lag}_1,\text{AveED}) + ti(\text{Lag}_2,\text{AveED}) + \xi_{t(DP)} \tag{4.5.5}$$

$$y_{t,DP} = \beta_0 + s_{1(DP)}(Lag_1) + s_{2(DP)}(Lag_2) + s_{3(DP)}(\text{DPED}) + s_{4(DP)}(\text{minED}) + s_{5(DP)}(\text{AveED}) + s_{6(DP)}(\text{AminT}) + s_{7(DP)}(\text{AveT}) + s_{8(DP)}(\text{AmaxT}) + s_{9(DP)}(\text{DBH}) + s_{10(DP)}(\text{DH}) + s_{11(DP)}(\text{DAH}) + s_{12(DP)}(\text{Month}) + s_{13(DP)}(\text{DayType}) + s_{14(DP)}(\text{trend}) + t2(\text{Lag}_1,\text{AveED}) + t2(\text{Lag}_2,\text{AveED}) + \xi_{t(DP)} \tag{4.5.6}$$

$$y_{t,DP} = \beta_0 + s_{1(DP)}(Lag_1) + s_{2(DP)}(Lag_2) + s_{3(DP)}(\text{DPED}) + s_{4(DP)}(\text{minED}) + s_{5(DP)}(\text{AveED}) + s_{6(DP)}(\text{AminT}) + s_{7(DP)}(\text{AveT}) + s_{8(DP)}(\text{AmaxT}) + s_{9(DP)}(\text{DBH}) + s_{10(DP)}(\text{DH}) + s_{11(DP)}(\text{DAH}) + s_{12(DP)}(\text{Month}) + s_{13(DP)}(\text{DayType}) + s_{14(DP)}(\text{trend}) + s_{15(DP)}(\text{Lag}_1,\text{AveED}) + s_{15(DP)}(\text{Lag}_2,\text{AveED}) + \xi_{t(DP)} \tag{4.5.7}$$

From models (4.5.4), (4.5.5), (4.5.6) and (4.5.7), the best model is selected using Adjusted R^2, Akaike information criterion (AIC), Bayesian Information Criterion (BIC) and Generalized Cross Validation (GCV). The smallest values of the AIC, BIC and GCV along with the highest value of Adjusted R^2 represent the best model. Since all models has same values of Adjusted R^2 and nearly similar values of AIC and BIC, we consider the model with

Table 4.16: Model selection for tensor product interactions

Model	Adjusted R^2	AIC	BIC	GCV
Model (4.5.4)	0.987	15348.36	15739.45	73714
Model (4.5.5)	0.987	15350.51	15732.91	73842
Model (4.5.6)	0.987	15353.34	15741.05	74045
Model (4.5.7)	0.987	15337.77	15804.33	73174

lowest value of GCV as best model. From Table 4.16, the best model is model (4.5.7). Figure 4.15 displays four standard diagnostic plots for the residuals, used for evaluating the fitted best model (4.5.7). From Figure 4.15, the top-left and bottom-left panels show that the residuals are seems to be normally distributed. Histogram also shows that the residuals is left (negatively) skewed. The top-left and bottom-right panels show the plots of the positive residual demand. The best model fit requires the residuals to be normally distributed and no pattern in the predicted values, as look like to be satisfied by model (4.5.7).

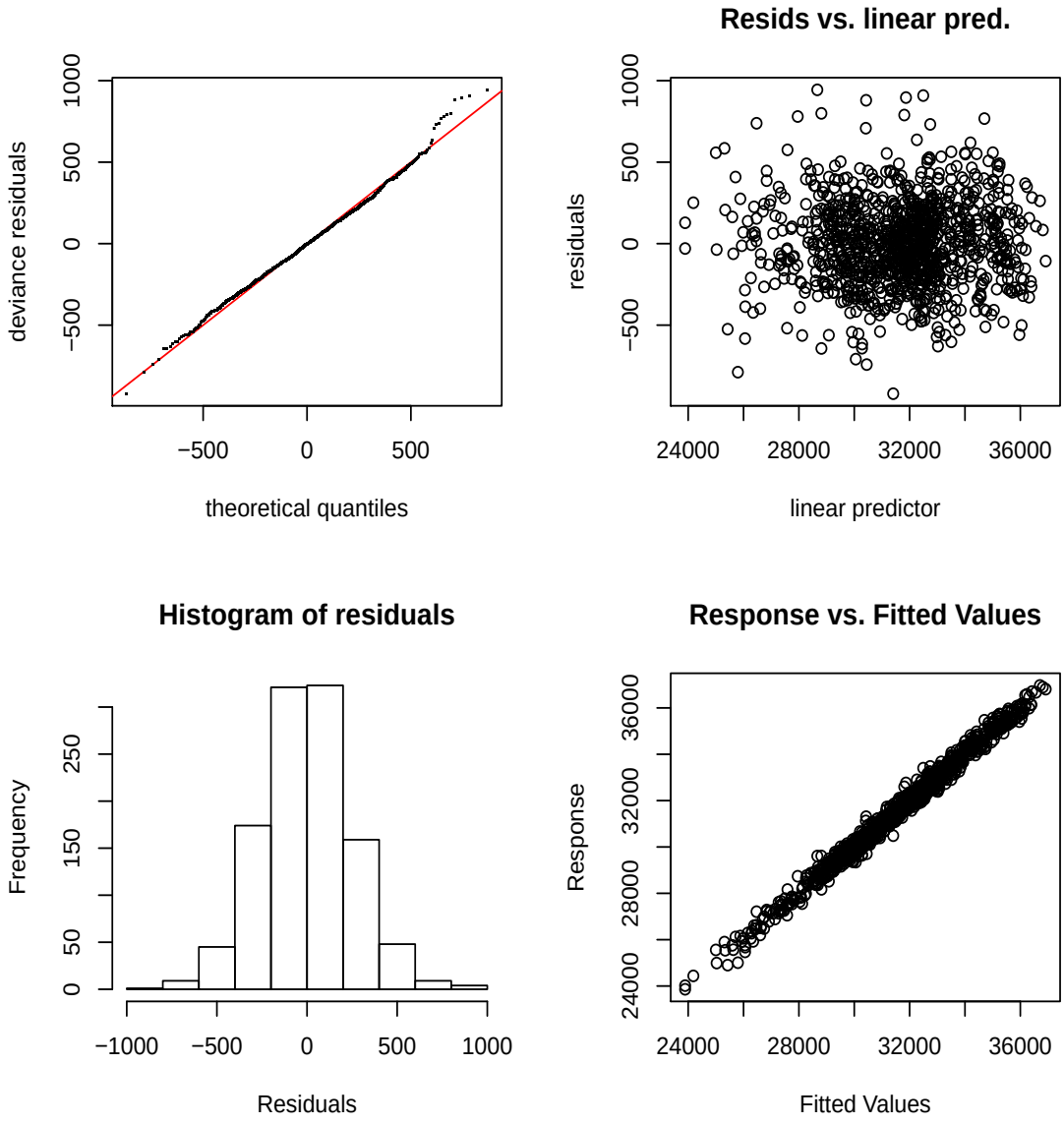

Figure 4.15: Diagnostic plots for the residuals for model (4.5.7). **Top left**: Quantile-Quantile (Q-Q) plot of residuals. **Top right**: linear predictor vs. residuals. **Bottom left**: histogram of residuals. **Bottom right**: the plot of fitted values vs. response.

4.5.6 Evaluation of forecasts from best models with and without tensor product interactions

MAPE, MAE and RMSE are used to evaluate the forecasting accuracy of best prediction models. We consider best model (4.5.3) without tensor product interactions and best model (4.5.7) with smoothing interactions. The model with the smaller MAPE, MAE and RMSE is the best model, which is model (4.5.7). Table 4.17 displays accuracy measures for evaluation of forecasts during testing period.

Table 4.17: Evaluation of forecasts from best models with and without tensor product interactions for DPED

Performance criteria	Model 4.5.3	Model 4.5.7
MAPE	27.95	26.82
MAE	6421.53	6244.60
RMSE	7111.34	6893.61

4.6 Comparing analysis of forecasts

In this section we compare forecasts using forecasts from maximum demand for peak hours 19:00 and 20:00, and the forecasts from daily peak electricity demand. The objective is to find the best way of predicting forecasts for peak electricity demand. Models which give smaller forecast accuracy measures are preferred to those with larger error measures. Table 4.18 displays accuracy measures for evaluation of forecasts using maximum forecasts from peak

hours and DPED during the testing period. Figure 4.16 displays accuracy

Table 4.18: Performance evaluation of forecasts for maximum forecasts of peak hours and forecasts of DPED.

Performance criteria	Maximum peak hour	DPED
MAPE	0.63	1.25
MAE	193.26	389.40
RMSE	250.86	489.48

measures for maximum peak hour and DPED during testing period.

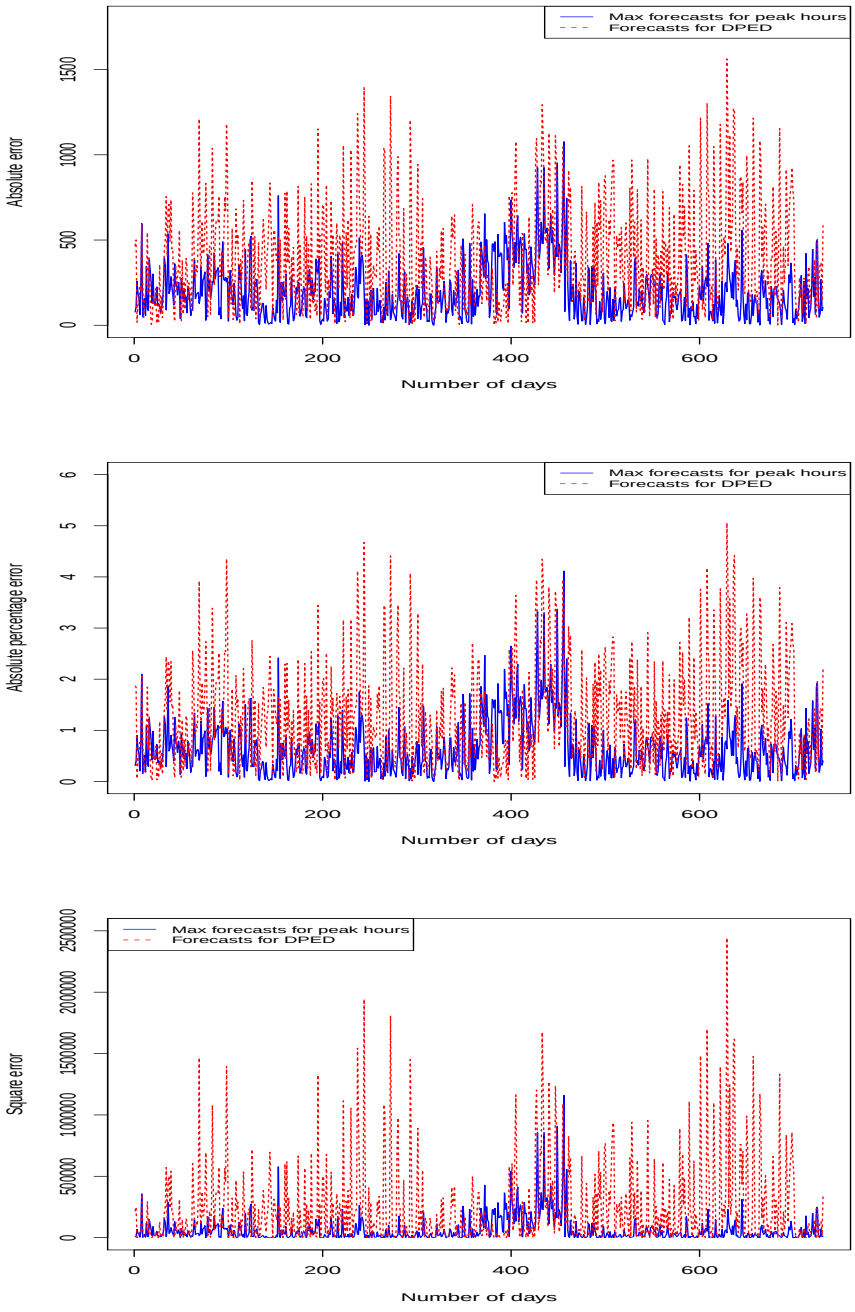

Figure 4.16: Accuracy measures for evaluation of forecasts for maximum peak hours and DPED. **Top panel**: Absolute error for testing period. **Middle panel**: Absolute percentage error for testing period. **Bottom panel**: Square error for testing period.

4.7 Operational forecasts

Operational forecasts are produced without access of predictors variables. However, the predictor variables DBH, DH, DAH, Month, DayType and trend are used to forecast the unknown predictor variables which are Lag_1, Lag_2, DPED, minED, AveED, AminT, AveT and AmaxT. The forecasted predictor variables are then used to forecast electricity demand at hour 19:00. Although in this study we are using hour 19:00, it should be noted that the developed modelling framework can be applied to any other hours of the day. Figure 4.17 shows that the forecasts follow the actual demand remarkably well.

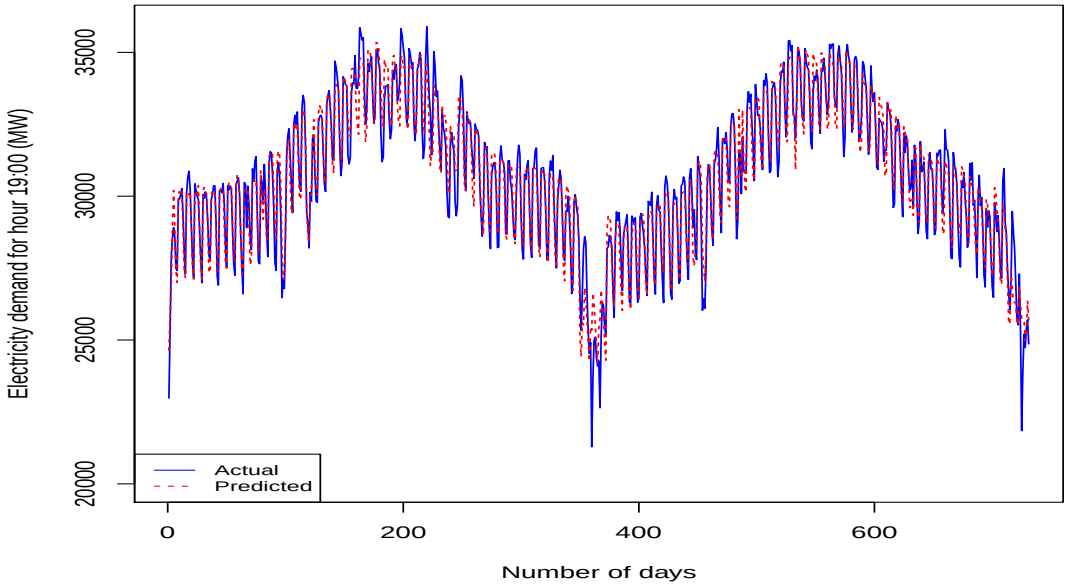

Figure 4.17: Actual and predicted electricity demand in megawatts at hour 19:00 from January 2012 to December 2013 (operational forecasts).

4.8 Performance comparison of best model at hour 19:00

4.8.1 Quantile regression averaging

This section presents the comparison of model performance using demand at 19:00. The study uses demand at hour 19:00, it should be noted that the demand at hour 20:00 and daily peak electricity demand can also be used also to compare model performance. In order to find the prediction intervals and averaging forecasts for proposed models from hour 19:00, we use Quantile regression averaging (QRA). The forecasts of proposed models are used as a

independent variables and demand at hour 19:00 are treated as dependent variable. The proposed best models are model (4.3.2) from Lasso, model (4.3.3) from Lasso via hierarchical interactions and model (4.3.6) from tensor product interaction. M1F, M2F and M3F denote model (4.3.2), model (4.3.3) and model (4.3.6) respectively. Table 4.19 presents the accuracy measures for proposed models of hour 19:00, before and after QRA. QRAM4 is based on QRA of M1F, M2F and M3F. QRAM4 is given by:

$$y_{t,\tau(QRA)} = \beta_0 + \beta_1 M1F + \beta_2 M2F + \beta_3 M3F + \xi_t \tag{4.8.1}$$

Based on the accuracy measures QRAM4 is considered to be the best model.

Table 4.19: Evaluation of forecasts from best models at hour 19:00 before and after QRA

Accuracy measures from best models at hour 19:00 before QRA				
Criteria	M1F	M2F	M3F	
MAPE	1.279	0.68	0.67	
MAE	381.609	204.58	200.07	
RMSE	481.788	263.38	261.05	
Accuracy measures from best models at hour 19:00 after QRA				
Criteria	QRAM1	QRAM2	QRAM3	QRAM4
MAPE	1.092	0.759	0.636	0.532
MAE	319.588	220.766	189.932	156.941
RMSE	470.18	339.984	251.257	228.253

4.8.2 Prediction interval

Table 4.20 shows that M1F, M2F and M3F have valid PICPs (PICP>95%). This means that more than 95% of the actual demands (targets) are inside the prediction intervals. PICP for all models are valid since they are greater than nominal confidence level of PIs (PICP>0.95). The densities on Figure 4.18 show that the distributions of the three models do not follow normal distributions. It can be seen also that the averaging model QRAM4 produces best forecasts compared to all proposed models. The PINAW of QRAM4 is 6.35% which is the narrowest PINAW. According to Sun et al. (2017) the model with the narrowest PINAW and smallest PINAD is considered to be the best fitting model.

Table 4.20: Comparisons of proposed model QRAM1, QRAM2, QRAM3 and QRAM4.

Model	PICP	PINAW	PINAD
QRAM1	96.58%	10.15%	0.02%
QRAM2	96%	7.56%	0.04%
QRAM3	95.35%	6.75%	0.04%
QRAM4	95.49%	6.35%	0.02%

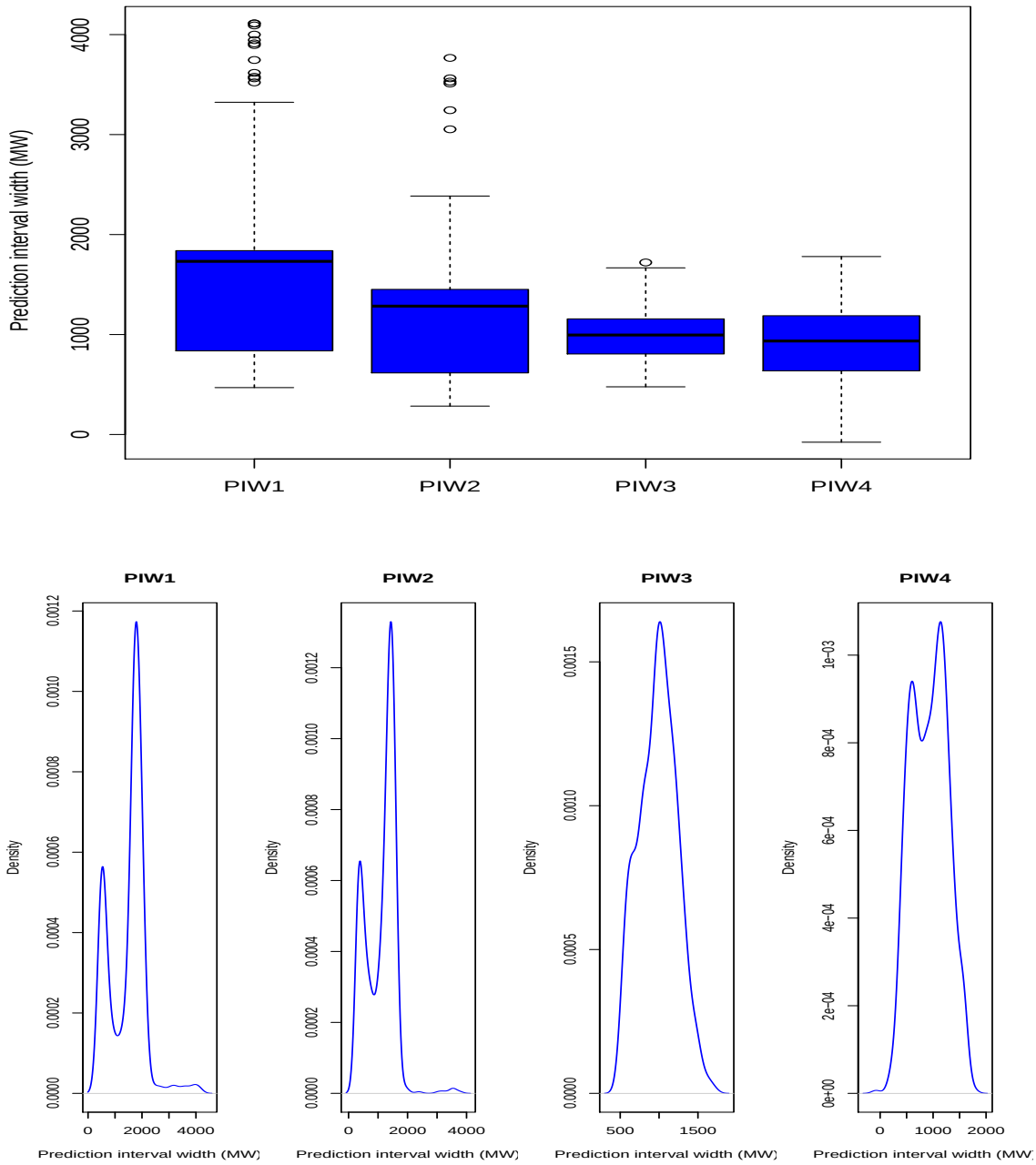

Figure 4.18: **Top panel**: Box-plots of widths of the prediction intervals (PI) for QRAM1, QRAM2, QRAM3 and QRAM4. **Bottom right**: Density plots of widths of the prediction intervals (PI) for model QRAM1, QRAM2, QRAM3 and QRAM4.

4.8.3 Forecast error distribution

The forecast error statistics for models at hour 19:00 are given in Table 4.21. The negative value of skewness for all models indicate a large number of negative errors. Negative skewness reflect many observations which are under-prediction. It can be seen in Figure 4.19 that the forecast errors for all models are negative skewed. The risk of using QRAM1 to predict forecasts is that it produces less forecasts than what is expected. A high kurtosis value greater than 3 can be interpreted as a failure to predict high load occurrences.

Table 4.21: Descriptive statistics of forecast errors for models QRAM1, QRAM2, QRAM3 and QRAM4

Hour	Min	Q_1	Q_2	Mean	Q_3	Max	Skew	Kurtosis	Under	Over
QRAM1	-3032	-250.5	1	-63.28	165	1157	-1.034	4.024	370	361
QRAM2	-2589	-143.5	-6	-58.47	119.5	1019	-1.628	7.055	356	375
QRAM3	-915	-154.5	1	-2.26	144.5	1044	-0.027	1.212	366	365
QRAM4	-863	-98.5	0	-2.12	93	1030	-0.075	2.207	366	365

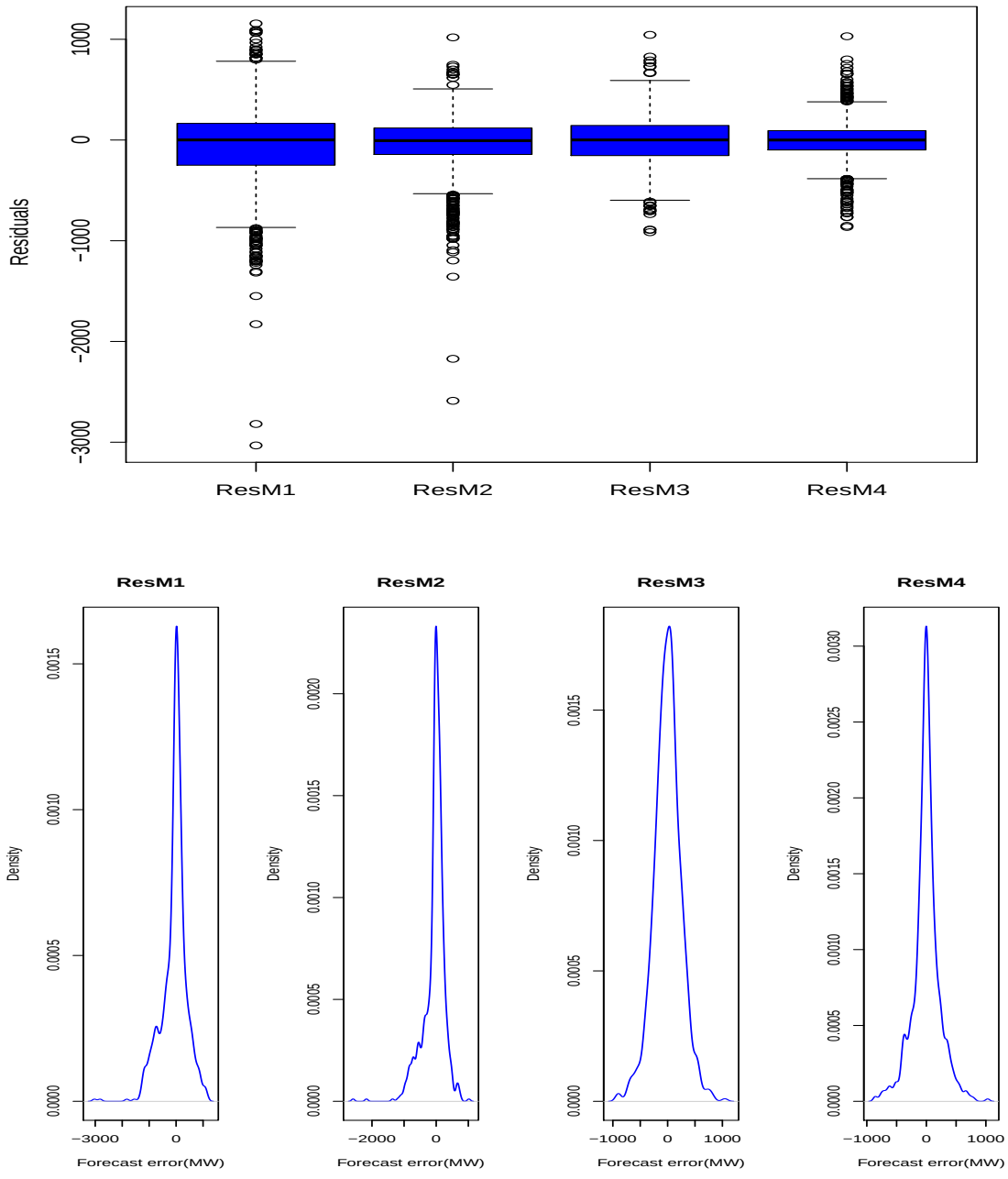

Figure 4.19: **Top panel**: Box-plots of forecasts errors for QRAM1, QRAM2, QRAM3 and QRAM4. **Bottom right**: Density plots of forecasts errors for models QRAM1, QRAM2, QRAM3 and QRAM4.

4.8.4 Percentage Improvement

Table 4.22 presents percentage improvement of the performance criterion using mean absolute error for best model obtained in Table 4.20. From Table 4.20 the best model is QRAM4, therefore percentage improvements are calculated for other models after QRA against the best model QRAM4 and presented in Table 4.22. The highest percentage improvement is for QRAM4 over QRAM1 which is 50.89% and the smallest percentage improvement is for QRAM4 over QRAM3 which is 17.37%. This also implies that QRAM3 is the second best model. The best model for prediction of electricity demand is QRAM4 followed by the QRAM3.

Table 4.22: Percentage improvement for hour 19:00.

Models	$MAE_{best\ model}$	$MAE_{other\ model}$	Improvement (%)
QRAM1&QRAM4	156.941	319.588	50.89%
QRAM2&QRAM4	156.941	220.766	28.91%
QRAM3&QRAM4	156.941	189.932	17.37%

4.9 Concluding Remarks

The chapter presented and discussed the modelling of daily peak electricity demand in different hours of the day using generalized additive models with tensor product interactions using South African hourly data of the period 1 Jan 2009 to 31 Dec 2013. Electricity demand from different hours were modelled in order to find the peak hour of the day. Most of the peak demands

occur in the hours 19:00 and 20:00. The chapter proposed generalized additive models and generalized additive models with tensor product interactions for forecasting demand at hours 19:00 and 20:00, and daily peak electricity demand. Operational forecasts are proposed in order to forecast the demand with unknown predictor variables. Comparing forecasts from different developed models at hour 19:00 were done using quantile regression averaging. The next chapter presents the conclusion of this study.

Chapter 5

Conclusion

5.1 Introduction

The study presented and analysed generalized additive models with tensor product interactions in forecasting medium term load in South Africa over the period 01 January 2009 to 31 December 2013. Medium term load forecasts are important in the planning of the operation of power utility companies such as Eskom, including medium term risk assessment. The study started by describing the data using frequency table and bar graph in order to find the frequency of occurrences of DPED in different hours of the day.

5.2 Summary and concluding remarks

In this section we give a brief summary of analyses that were discussed in Chapter 4. According to findings in section 4.2, daily peak electricity demand occurs on hours 19:00 and 20:00. The electricity demand is increasing at both hours, 19:00 and 20:00 from 2009 to 2010 and steadily decreasing from 2011

(see Figure 4.2). This information is important for decision-making which includes planning for future energy consumption.

Section 4.3 was based on the findings obtained in Section 4.2, generalized additive models with tensor product interactions were developed using electricity demand data for hour 19:00. Lasso and Lasso via hierarchical interactions were used for variable selection in order to avoid unnecessary predictors and over-fitting. The best model was selected based on smallest values of Akaike information criterion, Bayesian information criterion and Generalized cross validation along with highest values of Adjusted R^2. The findings from the testing periods confirmed that the generalized additive models with tensor product interactions give more accurate forecasts compared to generalized additive models without interactions based on the MAPE, MAE and RMSE. There is some evidence that forecasts from generalized additive models with tensor product interactions follow the actual demand remarkably well (see Figure 4.7).

Generalized additive models with tensor product interactions for hour 20:00 were developed in Section 4.4. Lasso and Lasso via hierarchical interactions were used for variable selection. The model based on Lasso via hierarchical interactions gave better forecasts. We also developed generalized additive models with tensor product interactions, treating DPED as a dependent variable. The model with tensor product interactions was found to be better for

forecasting DPED. Tensor product interactions were used to evaluate the relationship between interacting variables (DayType and Month) and demand at hours 19:00 and 20:00. The highest peak during the week days occurs from Monday to Thursday and is mostly observable within May to late July (see Figure 4.6 and 4.11). This is due to winter season and it has been established that demand for electricity in South Africa increases significantly during winter season and slightly during the summer season (Mokilane et al. 2018).

Findings on Section 4.6, showed that the best way of forecasting electricity using generalized additive models with tensor product interactions is when we forecasted from peak hours i.e 19:00 and 20:00 (see Table 4.18). Operational forecasts for hour 19:00 were also given. The forecasted demand from operational forecasting also shows that the forecasts follow the actual demand remarkably well (see Figure 4.17).

The performance of the proposed models for hour 19:00 were compared and generalized additive models with tensor product interactions model (4.3.6) is found to be the best fitting model. The forecast error (residual) distribution also confirmed that model (4.3.6) is the best model. There is an equal number of underestimation and overestimation predictions. The percentage improvement for averaging model over model (4.3.6) is 17.37% , which is smallest compared to the improvement of averaging model over other mod-

els. This information is useful for system operators to know when is the underestimation which can lead to power outages, load shedding, blackouts and ultimately threaten to destabilize the national grid. Although the study focused on the analysis of the models for hours 19:00 and 20:00 it is important to note that any hour can be done using same modelling approach.

5.3 Limitations of the dissertation

The limitations of this study is that only aggregated hourly electricity demand for all the sectors of the South African economy which are: residential, industrial, commercial and agricultural sectors was used.

5.4 Future Research

This study used aggregated national data. Future research should look at forecasting electricity demand for each of the nine provinces of South Africa separately including analysis by each sector (i.e. residential, commercial, agricultural and industrial sectors). Some provinces are in the coastal areas and electricity demand patterns are expected to be different from provinces which are inland. Some of the provinces such as Gauteng are densely populated and heavily industrialized as compared to others. The future study will include use of other forecasting techniques such as additive quantile regression, etc.

References

N. Abu-Shikhah, F. Elkarmi, and O. M. Aloquili. Medium-term electric load forecasting using multivariable linear and non-linear regression. *Smart Grid and Renewable Energy*, 2(02):126, 2011.

M. Abuella and B. Chowdhury. Hourly probabilistic forecasting of solar power. In *Power Symposium (NAPS), 2017 North American*, pages 1–5. IEEE, 2017.

H Akaike. Information theory and an extension of the maximum likelihood principal. In *2nd International Symposium on Information Theory, 1973*. Akademiai Kiado, 1973.

D.K. Barrow and N. Kourentzes. Distributions of forecasting errors of forecast combinations: Implications for inventory management. *International Journal of Production Economics*, 177:24–33, 2016.

K. Berk and A. Müller. Probabilistic forecasting of medium-term electricity demand: A comparison of time series models. 2016.

J. Bien, J. Taylor, and R. Tibshirani. A lasso for hierarchical interactions. *Annals of statistics*, 41(3):1111, 2013.

P. Bihani and S.T. Patil. A comparative study of data analysis techniques. *International Journal of Emerging Trends & Technology in Computer Science*, 3(2):95–101, 2014.

F. Cavallaro. Electric load analysis using an artificial neural network. *International journal of energy research*, 29(5):377–392, 2005.

T. Chai and R. R. Draxler. Root mean square error (rmse) or mean absolute error (mae)?–arguments against avoiding rmse in the literature. *Geoscientific Model Development*, 7(3):1247–1250, 2014.

H. Cho, Y. Goude, X. Brossat, and Q. Yao. Modeling and forecasting daily electricity load curves: a hybrid approach. *Journal of the American Statistical Association*, 108(501):7–21, 2013.

A. Chouldechova and T. Hastie. Generalized additive model selection. *arXiv preprint arXiv:1506.03850*, 2015.

R. D. Cook and S. Weisberg. *Residuals and influence in regression*. New York: Chapman and Hall, 1982.

P. Craven and G. Wahba. Smoothing noisy data with spline functions. estimating the correct degree of smoothing by the method of generalized cross-validation. *Numerische Mathematik*, 31:317–403, 1979.

E. A. Feinberg and D. Genethliou. Load forecasting. In *Applied mathematics for restructured electric power systems*, pages 269–285. Springer, 2005.

V. Fonti and E. Belitser. Feature selection using lasso, 2017.

M. Ghiassi, D. K. Zimbra, and H. Saidane. Medium term system load forecasting with a dynamic artificial neural network model. *Electric Power Systems Research*, 76(5):302–316, 2006.

Y. Goude, R. Nedellec, and N. Kong. Local short and middle term electricity load forecasting with semi-parametric additive models. *IEEE transactions on smart grid*, 5(1):440–446, 2014.

C. Gu. 1994 wald memorial lectures-polynomial splines and their tensor products in extended linear modeling-discussion, 1997.

T. Hastie and M. T. Hastie. Package gam, 2017.

T. J. Hastie and R. J. Tibshirani. *Generalized additive models*, volume 43. CRC press, 1990.

T. Hong and S. Fan. Probabilistic electric load forecasting: A tutorial review. *International Journal of Forecasting*, 32(3):914–938, 2016.

T. Hong, P. Pinson, S. Fan, H. Zareipour, A. Troccoli, and R. J. Hyndman. Probabilistic energy forecasting: Global energy forecasting competition 2014 and beyond, 2016.

R. J. Hyndman and S. Fan. Density forecasting for long-term peak electricity demand. *IEEE Transactions on Power Systems*, 25(2):1142–1153, 2010.

R.J. Hyndman and G. Athanasopoulos. *Forecasting: principles and practice*. OTexts, 2014.

G. James, D. Witten, T. Hastie, and R. Tibshirani. *An introduction to statistical learning*, volume 112. Springer, 2013.

P. Laurinec. Doing magic and analyzing seasonal time series with gam in r. Time series data mining in R. Bratislava, Slovakia, 2017.

M. Lim and T. Hastie. Learning interactions via hierarchical group-lasso regularization. *Journal of Computational and Graphical Statistics*, 24(3): 627–654, 2015.

B. Liu, J. Nowotarski, T. Hong, and R. Weron. Probabilistic load forecasting via quantile regression averaging on sister forecasts. *IEEE Transactions on Smart Grid*, 8(2):730–737, 2017.

H. Liu. Generalized additive model. *Department of Mathematics and Statistics, University of Minnesota Duluth, Duluth, MN*, 55812, 2008.

M. Mohandes. Support vector machines for short-term electrical load forecasting. *International Journal of Energy Research*, 26(4):335–345, 2002.

P. Mokilane, J. Galpin, Sarma V.S. Yadavalli, P. Debba, R. Koen, and S. Sibiya. Density forecasting for long-term electricity demand in south africa using quantile regression. *South African Journal of Economic and Management Sciences*, 21(1):14, 2018.

T. Nakata. Energy-economic models and the environment. *Progress in Energy and Combustion Science*, 30(4):417–475, 2004.

N. M. Odhiambo. Electricity consumption and economic growth in south africa: A trivariate causality test. *Energy Economics*, 31(5):635–640, 2009.

N. M. Odhiambo. Energy consumption, prices and economic growth in three SSA countries: A comparative study. *Energy Policy*, 38(5):2463–2469, 2010.

A. Pierrot and Y. Goude. Short-term electricity load forecasting with generalized additive models. *Proceedings of ISAP power*, 2011, 2011.

N. Pya and S. N. Wood. A note on basis dimension selection in generalized additive modelling. *arXiv preprint arXiv:1602.06696*, 2016.

H. Quan, D. Srinivasan, and A. Khosravi. Particle swarm optimization for construction of neural network-based prediction intervals. *Neurocomputing*, 127:172–180, 2014.

S. M. Ross. *Introduction to probability and statistics for engineers and scientists*. Academic Press, 2014.

I. A. Samuel, F. N. Chihurumanya, A. Adeyinka, and A. A. Awelewa. Medium-term load forecasting of Covenant University using the regression analysis methods. *Journal of Energy Technologies and Policy*, 4(4), 2014.

G. Schwarz. Estimating the dimension of a model. *The Annals of Statistics*, 6(2):461–464, 1978.

W. R. Shadish, A. F. Zuur, and K. J. Sullivan. Using generalized additive (mixed) models to analyze single case designs. *Journal of school psychology*, 52(2):149–178, 2014.

Y. Shen, X. Wang, and J. Chen. Wind power forecasting using multi-objective evolutionary algorithms for wavelet neural network-optimized prediction intervals. *Applied Sciences*, 8(2):185, 2018.

G. Singh, D. S. Chauhan, A. Chandel, D. Parashar, and G. Sharma. Factor affecting elements and short term load forecasting based on multiple linear

regression method. In *International Journal of Engineering Research and Technology*, volume 3. IJERT, 2014.

P. Smolensky. Tensor product variable binding and the representation of symbolic structures in connectionist systems. *Artificial intelligence*, 46 (1-2):159–216, 1990.

L. Suganthi and A. Samuel. Energy models for demand forecastinga review. *Renewable and sustainable energy reviews*, 16(2):1223–1240, 2012.

X. Sun, Z. Wang, and J. Hu. Prediction interval construction for byproduct gas flow forecasting using optimized twin extreme learning machine. *Mathematical Problems in Engineering*, 2017, 2017.

R. Tibshirani. Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 267–288, 1996.

L. Wasserman. Bayesian model selection and model averaging. *Journal of mathematical psychology*, 44(1):92–107, 2000.

S. Wood and M. S. Wood. Package mgcv. *R package version*, pages 1–7, 2017.

S. N. Wood. *Generalized additive models: an introduction with R*. London: Chapman and Hall, 2006.

S. N. Wood. *Generalized additive models: an introduction with R*. CRC press, 2017.

J. Xie, T. Hong, and C. Kang. From high-resolution data to high-resolution probabilistic load forecasts. In *Transmission and Distribution Conference and Exposition (T&D), 2016 IEEE/PES*, pages 1–5. IEEE, 2016.

Appendices

Appendix A

Some selected R codes

```
library(opera)
library(forecast)
library(tseries)
library(data.table)
library(mgcv)
library(car)
library(ggplot2)
library(grid)
library(animation)
library(glmnet)
library(hierNet)
library(e1071)
library(qgam)

H <- −read_excel("C : /Users/RaveleThakhani/Desktop/H.xlsx")
attach(H)
names(H)
head(H)
tail(H)
win.graph()
par(mfrow=c(2,2))
y=ts(H19,start = 2009,frequency = 365)
plot(y,main="Demand at 19:00 (2009-2013",xlab = "Year",ylab="Electricity
```

demand at 19:00 (MW)",col="blue")
DPED=density(H19)
plot(DPED,main="Density plot at 19:00 (2009-2013)",xlab = "Electricity demand (MW)",col="blue")

y=ts(H20,start = 2009,frequency = 365)
plot(y,main="Demand at 20:00 (2009-2013)",xlab = "Year",ylab="Electricity demand at 20:00 (MW)",col="blue")
DPED=density(H20)
plot(DPED,main="Density plot at 20:00 (2009-2013)",xlab = "Electricity demand (MW)",col="blue")

#Creating lags for 19H00

lag1< −diff(H19,lag=1)
length(lag1)
write.table(lag1,"∼/lag1.txt",sep="\t")
lag2< −diff(H19,lag=2)
length(lag2)
write.table(lag2,"∼/lag2.txt",sep="\t")

#Importing data with lags

attach(HL)
head(HL)
tail(HL)

#creating train and test

testing < − 1094:nrow(HL)

train < − HL[-testing,]

test < − HL[testing,]

length(test)

Generalized additive models

Variable selection Via LASSO for 19H00

y< −train$H19

x< −data.matrix(train[4:17])

print(x)

LAsso< −glmnet(y=y,x=x,family=“gaussian”)

plot(LAsso)

coef(LAsso)

cv.lasso< −cv.glmnet(y=y,x=x,family=“gaussian”)

cv.lasso

plot(cv.lasso)

coef(cv.lasso)

predict.1< −predict(cv.lasso,newx=x)

predict.1

m.lasso< −mean((y-predict.1)∧2)

m.lasso

#LASSO model for for 19H00

M1 < − gam(H19∼ s(DPED) + s(minED) + s(AveT) + s(AmaxT) + as.factor(DBH) + s(trend), data = train)

summary(M1)

#residual for 19H00

r1< −residuals.gam(M1)

r1

Box.test(r1, lag=20, fitdf=1, type="Ljung")

win.graph()

tsdisplay(r1,main = "(a) Residual Demand", col="blue")

auto.arima(r1)

f1< −Arima(r1,order = c(5,0,2),seasonal = list(order = c(2,0,3),periods=7))

#correcting autocorrelation for 19H00

win.graph()

tsdisplay(residuals(f1),main = "(b) Residual Demand", col="blue")

Box.test(residuals(f1), lag=20, fitdf=1, type="Ljung")

y3< −forecast(f1,h=12)

plot(y3,col = "blue")

Variable selection using Lasso via hierarchical interactions for 19:00

y< −train$H19

x< −data.matrix(train[4:17])

print(x)

#a typical analysis including cross-validation

fitH19=hierNet.path(x,y, strong=TRUE)

fitcvH19=hierNet.cv(fitH19,x,y)

print(fitcvH19)

plot(fitcvH19)

lamhat=fitcvH19$lamhat.1se

lamhat

fit1H19=hierNet(x,y,lam=lamhat, strong=TRUE)

print(fit1H19)

yhat=predict.hierNet(fit1H19,x)

yhat

plot(yhat)

lines(DPED,col="red")

#hiernet model for for 19H00

M2 < − gam(H19∼s(Lag1H19) + s(Lag2H19)+ s(DPED) + s(minED) + s(AveED) + s(AminT) + s(AveT) + s(AmaxT) + as.factor(DBH) + as.factor(DH) + as.factor(DAH) + as.factor(Month) + as.factor(DayType) + s(trend) + s(a) + s(b) + s(c) + s(d), data = train)

summary(M2)

#residual for 19H00 via hiernet

r2< −residuals.gam(M2)

r2

Box.test(r2, lag=12, fitdf=1, type="Ljung")

win.graph()

tsdisplay(r2,main = "(a) Residual Demand", col="blue")

auto.arima(r2)

f2< −Arima(r2,order = c(5,0,2),seasonal = list(order = c(2,0,3),periods=7))

#correcting autocorrelation

win.graph()

tsdisplay(residuals(f2),main = "(b) Residual Demand", col="blue")

Box.test(residuals(f2), lag=12, fitdf=1, type="Ljung")

y2< −forecast(f2,h=12)

plot(y2,col = "blue")

#model selection

AIC(M1,M2)

BIC(M2,M2)

Generalized additive models with tensor product interactions

te()

M3 < − gam(H19 ∼s(Lag1H19) + s(Lag2H19)+ s(DPED) + s(minED) + s(AveED) + s(AminT) + s(AveT) + s(AmaxT) + as.factor(DBH) + as.factor(DH) + as.factor(DAH) + as.factor(Month) + as.factor(DayType)+ s(trend) + te(Lag1H19, DPED) + te(Lag1H19, AmaxT) + te(Lag2H19, DPED) + te(Lag2H19, AmaxT), data = train)

```
# ti()
M4 < − gam(H19 ∼s(Lag1H19) + s(Lag2H19)+ s(DPED) + s(minED) +
s(AveED) + s(AminT) + s(AveT) + s(AmaxT) +
as.factor(DBH) + as.factor(DH) + as.factor(DAH) + as.factor(Month) +
as.factor(DayType)+ s(trend) +
ti(Lag1H19, DPED) + ti(Lag1H19, AmaxT) + ti(Lag2H19, DPED) + ti(Lag2H19,
AmaxT), data = train)

# t2()
M5 < − gam(H19 ∼s(Lag1H19) + s(Lag2H19)+ s(DPED) + s(minED) +
s(AveED) + s(AminT) + s(AveT) + s(AmaxT) +
as.factor(DBH) + as.factor(DH) + as.factor(DAH) + as.factor(Month) +
as.factor(DayType)+ s(trend) +
t2(Lag1H19, DPED) + t2(Lag1H19, AmaxT) + t2(Lag2H19, DPED) +
t2(Lag2H19, AmaxT), data = train)

# smoohing
Ms < − gam(H19 ∼s(Lag1H19) + s(Lag2H19)+ s(DPED) + s(minED) +
s(AveED) + s(AminT) + s(AveT) + s(AmaxT) +
as.factor(DBH) + as.factor(DH) + as.factor(DAH) + as.factor(Month) +
as.factor(DayType)+ s(trend) +
s(Lag1H19, DPED) + s(Lag1H19, AmaxT) + s(Lag2H19, DPED) + s(Lag2H19,
AmaxT), data = train)

summary(M3)
```

summary(M4)

summary(M5)

summary(Ms)

AIC(M3,M4,M5,Ms)

BIC(M3,M4,M5,Ms)

#Diagnostic plots for the residuals for model

win.graph()

par(mfrow=c(2,2))

gam.check(M4)

#Contour plot for model

win.graph()

par(mfrow=c(4,4))

plot(M4, rug = FALSE, se = FALSE, n2 = 80, main = "gam with ti()")

#3D view

win.graph()

M6 < − gam(H19 ∼ s(DayType, bs = "ps", k=7) + s(Month, bs = "ps") +
ti(DayType, Month,
k = c(7,7), bs = c("ps","ps"), fx = TRUE),
data = train,
family = gaussian)

vis.gam(M6, n.grid = 50, theta = 35, phi = 32, zlab = "Electricity demand

for hour 19:00 (MW)",
ticktype = "detailed", color = "topo", main = "(a) ti(Month, DayType)")

#2D contour

win.graph()
vis.gam(M6, main = "(b) ti(DayType, Month)", plot.type = "contour",
color = "terrain", contour.col = "black"", lwd = 2)

Comparing forecasts

#Importing forecasted data

attach(LassoP)
head(LassoP)
tail(LassoP)

#evaluation of forecasts for H19 for tensor and without tensor

accuracy(H19,LASSO)
accuracy(H19,H19Hier)
accuracy(H19,ti)

#Actual and predicted plots for 19H00 using ti()

win.graph()
u=ts(H19)
plot(u,xlab="Number of days",ylab="Electricity demand for hour 19:00 (MW)",
ylim=c(20000,36000),col="blue")

t1=ts(ti)

lines(t1,col=“red”,lty=2)

legend(“bottomleft”,legend = c(“Actual”,“Predicted”),col = c(“blue”,“red”),lty = 1:2,cex = 0.8)

#Actual and predicted density plots for 19H00

win.graph()

plot(density(H19),main = “Actual and predicted demand density at 19:00”,xlab = “Electricity demand (MW)”,col=“blue”)

lines(density(ti),col=“red”,lty=2)

legend(“topright”,legend=c(“Demand”,“Predicted”),col=c(“blue”,“red”),lty=1:2,cex=0.8)

Quantile regression averaging

 # Calibrate learning rate on a grid 0.025

set.seed(5235)

tun < − tuneLearnFast(form=H19∼s(M1Lasso, bs=“ad”),

err = 0.05, qu = 0.025, data = M1)

tun

fit1 < −qgam(H19∼s(M1Lasso, bs=“ad”),

err = 0.05, qu = 0.025, lsig = tun$lsig, data = M1)# insample

summary(fit1, se=“boot”) #se =“ ker” “ nid” “boot”

lines(fit1$fit, col=“red”)

```
#plot(fit1, pages=1)
#check.qgam(fit1, pch=19, cex=.3)

fQRAM1 = fitted(fit1)
write.table(fQRAM1,"~/fQRAM1025.txt",sep="\t")
# 0.5

# Calibrate learning rate on a grid
set.seed(5235)
tun <- tuneLearnFast(form=H19~s(M1Lasso, bs="ad"),
err = 0.05, qu = 0.5, data = M1)
tun
fit1 <-qgam(H19~s(M1Lasso, bs="ad"),
err = 0.05, qu = 0.5, lsig = tun$lsig, data = M1)# insample

summary(fit1, se="boot") #se =" ker" " nid" "boot"

lines(fit1$fit, col="red")

#plot(fit1, pages=1)
#check.qgam(fit1, pch=19, cex=.3)

fQRAM2 = fitted(fit1)
write.table(fQRAM2,"~/fQRAM105.txt",sep="\t")

# Calibrate learning rate on a grid 0.975
```

```
set.seed(5235)
tun < − tuneLearnFast(form=H19∼s(M1Lasso, bs=“ad”),
err = 0.05, qu = 0.975, data = M1)
tun
fit1 < −qgam(H19∼s(M1Lasso, bs=“ad”),
err = 0.05, qu = 0.975, lsig = tun$lsig, data = M1)# insample

summary(fit1, se=“boot”) #se =“ ker” “ nid” “boot”

lines(fit1$fit, col=“red”)

#plot(fit1, pages=1)
#check.qgam(fit1, pch=19, cex=.3)

fQRAM3 = fitted(fit1)
write.table(fQRAM3,“∼/fQRAM1975.txt”,sep=“\t”)
```

#prediction interval width

```
attach(AveM)
head(AveM)

win.graph()
PIW < − c(“PIW1”, “PIW2”, “PIW3”, “PIW4”)
boxplot(PIW1, PIW2, PIW3, PIW4, names= PIW, horizontal = FALSE,
main=“”,ylab=“Prediction interval width (MW)”, col = “blue”)
```

```
par(mfrow=c(1,4))
plot(density(PIW1),xlab="Prediction interval width (MW)", col="blue", main="PIW1")
plot(density(PIW2),xlab="Prediction interval width (MW)", col="blue", main="PIW2")
plot(density(PIW3),xlab="Prediction interval width (MW)", col="blue", main="PIW3")
plot(density(PIW4),xlab="Prediction interval width (MW)", col="blue", main="PIW4")

#Residual Error
summary(ResM1)
summary(ResM2)
summary(ResM3)
summary(ResM4)

#Skewness of residual
skewness(ResM1)
skewness(ResM2)
skewness(ResM3)
skewness(ResM4)

#Kurtosis of residual
kurtosis(ResM1)
kurtosis(ResM2)
kurtosis(ResM3)
kurtosis(ResM4)

#Residual box-pot and density plot
```

```
RESID <- c("ResM1", "ResM2", "ResM3", "ResM4")
boxplot(ResM1, ResM2, ResM3, ResM4, names= RESID, horizontal = FALSE,
main="",ylab=" Residuals", col = "blue")

par(mfrow=c(1,4))
plot(density(ResM1),xlab="Forecast error(MW)", col="blue", main="ResM1")
plot(density(ResM2),xlab="Forecast error(MW)", col="blue", main="ResM2")
plot(density(ResM3),xlab="Forecast error(MW)", col="blue", main="ResM3")
plot(density(ResM4),xlab="Forecast error(MW)", col="blue", main="ResM4")
```

Appendix B

Plots

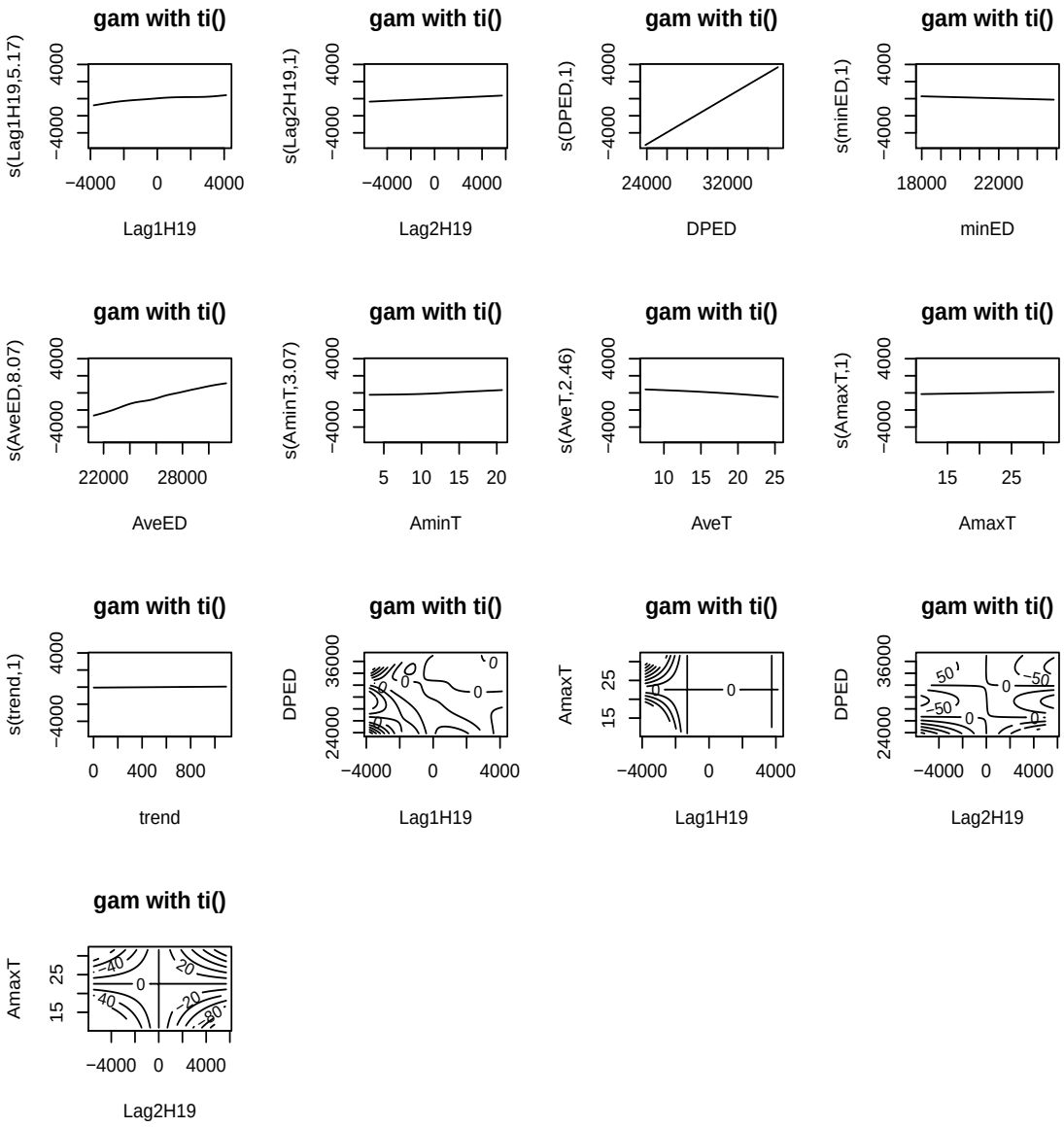

Figure 1: Contour plot for model (4.3.6).

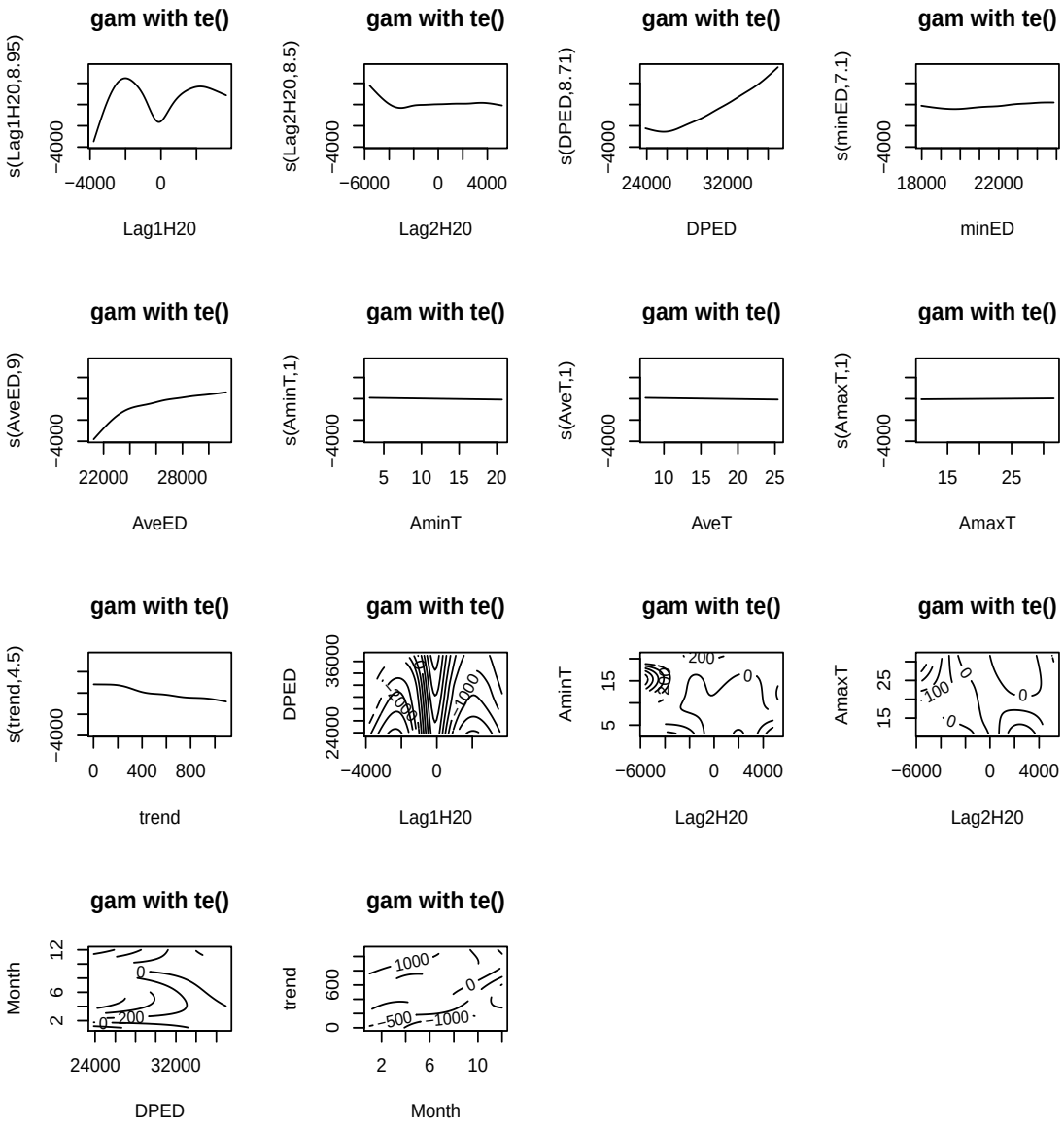

Figure 2: Contour plot for model (4.4.4).

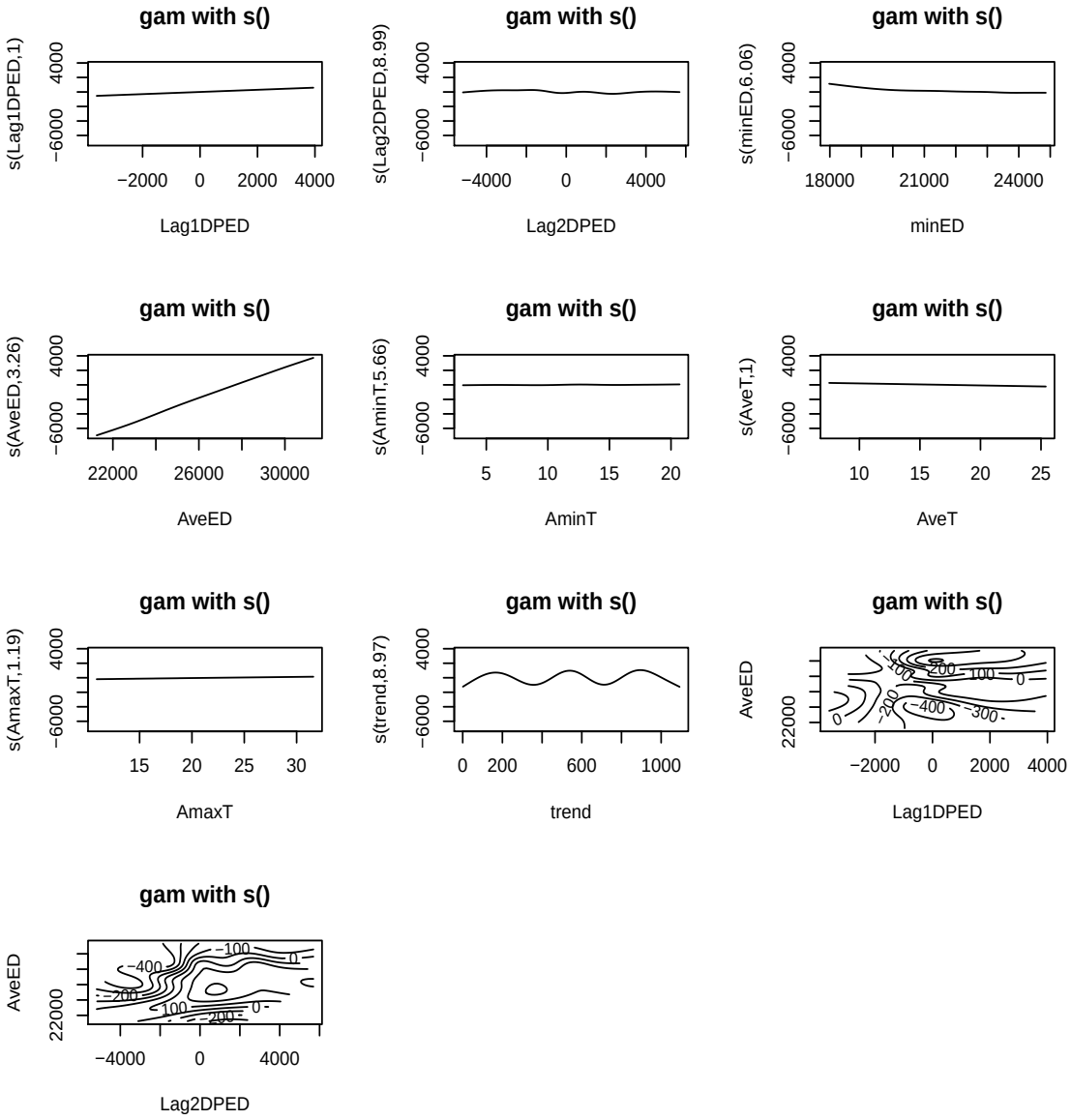

Figure 3: Contour plot for model (4.5.7)